

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2002, 04:52:54 : Search time 63.01 Seconds  
(without alignments)  
824.989 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVPSLVNGYDVATMAA.....IEAGNLLNVAKKAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2363	100.0	468	22	AAB66707 C-glutamicum phosph
2	2363	100.0	661	22	AAAG94650 C glutamicum prote
3	2332	98.7	661	22	AAB69080 Brevibacterium lac
4	1821	77.1	362	22	AAB66708 C-glutamicum phosph
5	660	27.9	627	22	AAU37874 Streptococcus pneu
6	480.5	20.3	704	22	AAU60962 Propionibacterium
7	468.5	19.8	683	22	AAU93207 C-glutamicum prote
8	468.5	19.8	683	22	AAB66721 C-glutamicum phosph
9	345.5	14.6	679	22	AAU34162 Staphylococcus aur
10	345.5	14.6	681	22	AAU36677 Staphylococcus aur
11	338.5	14.3	484	22	AAU35216 Enterococcus faeca

12	323	13.7	484	22	AAU36679 Staphylococcus aur
13	305.5	12.9	455	22	AAU38201 Salmonella typhi c
14	300	12.7	687	22	AAU34326 Staphylococcus aur
15	300	12.7	719	22	AAU37349 Staphylococcus aur
16	298.5	12.6	648	22	AAU34496 E. coli cellular p
17	298.5	12.6	648	22	AAU34496 Escherichia coli p
18	292	12.4	551	22	AAU34907 Enterococcus faeca
19	288.5	12.2	439	22	AAU34164 Staphylococcus aur
20	250.5	10.6	474	22	AAU34650 E. coli cellular p
21	247	10.5	523	22	AAU46075 Propionibacterium
22	234	9.9	135	22	AAU66723 C-glutamicum phosph
23	233.5	9.9	167	20	AAU20012 B. burgdorferi ant
24	232.5	9.8	583	21	AAU53620 Amino acid sequenc
25	229	9.7	280	22	AAU36886 Staphylococcus aur
26	228	9.6	205	22	AAU66872 Propionibacterium
27	228	9.6	207	22	AAU41453 Propionibacterium
28	226.5	9.6	842	22	AAU36357 Pseudomonas aerugi
29	221.5	9.4	263	22	AAU37425 Staphylococcus aur
30	221.5	9.4	263	22	AAU37559 Staphylococcus aur
31	220.5	9.3	135	22	AAU34346 Staphylococcus aur
32	201.5	8.5	142	22	AAU34222 Staphylococcus aur
33	197	8.3	168	22	AAU54924 Propionibacterium
34	189.5	8.0	173	22	ABG14977 Novel human diagno
35	177.5	7.5	285	22	ABG14978 Novel human diagno
36	168.5	7.1	118	22	AAU52635 Propionibacterium
37	166.5	7.0	135	20	AAU20013 B. burgdorferi ant
38	166.5	7.0	413	22	AAU66722 C-glutamicum phosph
39	165	7.0	640	22	AAU35193 Enterococcus faeca
40	161.5	6.8	126	18	AAU28051 Amino acid sequenc
41	156	6.6	449	21	AAU81556 Streptococcus pneu
42	154	6.5	108	21	AAU40791 Human ORFX ORF555
43	151.5	6.4	558	21	AAU15995 E. coli proliferat
44	149.5	6.3	375	22	ABG24413 Novel human diagno
45	141.5	6.0	639	22	ABB52613 Escherichia coli p

ALIGNMENTS

RESULT 1  
AAB66707  
ID AAB66707 standard; protein; 468 AA.  
XX  
AC AAB66707;  
XX  
DT 09-APR-2001 (first entry)  
DE  
DE C-glutamicum phosphoenolpyruvate protein #1.  
XX  
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200102583-A2.  
XX  
PD 11-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-IB00973.  
XX  
PR 01-JUL-1999; 99US-0142691.  
PR 23-AUG-1999; 99US-0150310.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042097.  
XX  
(BADI ) BASF AG.  
XX  
PI Pompeius M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX  
DR WPI; 2001-080989/09.  
XX  
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;  
PT sugar phosphotransferase system proteins or their portions, useful for  
PT typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -

PS Claim 4; Page 101-102; 144pp; English.

XX The present invention relates to Corynebacterium glutamicum  
 CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.  
 CC The PTS nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or  
 CC related bacteria, the typing or identification of C. glutamicum or  
 CC related bacteria, as reference points for mapping C. glutamicum genome,  
 CC and as markers for transformation.

XX Sequence 468 AA;

Query Match 100.0%; Score 2363; DB 22; Length 468;

Best Local Similarity 100.0%; Pred. No. 3e-227;  
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVFPSSLVNGYDVAAATMAAGEMPWMSLFGLDVAQAGYQGVLPVLVSVWILLATIEKFLH 60  
 DB 1 mamvfpslvngydvaaatmaagempwmslfgldvaqagvggtvlpvlvsvwillatiekflh 60  
 QY 61 KRLKGTADFLITPVLTLTLTGFTLFTIAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFLGL 120  
 DB 61 krlkgtdaflitpvltltlgtftlftiaigpamrwvgdvlaahglgglgldfpgpgvllflgl 120  
 QY 121 VYSPVITGLHQSPPIELFNGGSGFIFATASMANIAQGAACIATVFLAKSEKLGKGLA 180  
 DB 121 vyspviitglhqsppielelfnngsgsfifatasmanlaqgaaciatvflakseklkgl 180  
 QY 181 GASGVSAVLGTETPAIFGVNLRWPFIFGTAAIGGALLIALFNKAVAGAGFLGV 240  
 DB 181 gasgvsavlgtepaifgvnlrlwppifgttaagallialfnkavagagflgv 240  
 QY 241 SDAPDMVMFLVCVAVTFFIAFGAAIAYGLVLRNGSIDPDATAAPVPACTTAAEAP 300  
 DB 241 sidapdmvmflvcavtffiafgaaiaayglvlrngsidpdataapvpaacttaaeap 300  
 QY 301 AEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIIVAFP 360  
 DB 301 aefnsdstiiqapltgeaialssvsdamfasklgsgvaiivptkgqlvspvsgkiivafp 360  
 QY 361 SGHAFVTRTKAEDGSNDVILMHIGFDTVNLNGTHFNPLKKGDEVKAGELICEFDIDA 420  
 DB 361 sghafavtrtkaedgsndvilmhigfdtnlngthfnplkkgdevkagelicefdidaik 420  
 QY 421 AAGYEVTTPIVVSNNYKKTGPVNTYGLGIEAGANLLNVAKKEAVPATP 468  
 DB 421 aagyevttpivvsnnykktgpvntyglgieaganllnvakkeavpatp 468

RESULT 2

AAG92650

ID AAG92650 standard; Protein; 661 AA.

XX AAG92650;

XX 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6404.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

DR N-PSDB; AAH67869.

XX

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 XX -

PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.

XX

CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium glutamicum glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from corynebacterium, and identifying a homologue of a gene derived  
 CC from corynebacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 661 AA;

Query Match 100.0%; Score 2363; DB 22; Length 661;

Best Local Similarity 100.0%; Pred. No. 5e-227;

Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVFPSSLVNGYDVAAATMAAGEMPWMSLFGLDVAQAGYQGVLPVLVSVWILLATIEKFLH 60

DB 194 mamvfpslvngydvaaatmaagempwmslfgldvaqagvggtvlpvlvsvwillatiekflh 253

QY 61 KRLKGTADFLITPVLTLTLTGFTLFTIAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFLGL 120

DB 254 krlkgtdaflitpvltltlgtftlftiaigpamrwvgdvlaahglgglgldfpgpgvllflgl 313

QY 121 VYSPVITGLHQSPPIELFNGGSGFIFATASMANIAQGAACIATVFLAKSEKLGKGLA 180

DB 314 vyspviitglhqsppielelfnngsgsfifatasmanlaqgaaciatvflakseklkgl 373

QY 181 GASGVSAVLGTETPAIFGVNLRWPFIFGTAAIGGALLIALFNKAVAGAGFLGV 240

DB 374 gasgvsavlgtepaifgvnlrlwppifgttaagallialfnkavagagflgv 433

QY 241 SIDAPDMVMFLVCVAVTFFIAFGAAIAYGLVLRNGSIDPDATAAPVPACTTAAEAP 300

DB 434 sidapdmvmflvcavtffiafgaaiaayglvlrngsidpdataapvpaacttaaeap 493

QY 301 AEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIIVAFP 360

DB 494 aefnsdstiiqapltgeaialssvsdamfasklgsgvaiivptkgqlvspvsgkiivafp 553

QY 361 SGHAFVTRTKAEDGSNDVILMHIGFDTVNLNGTHFNPLKKGDEVKAGELICEFDIDA 420

DB 554 sghafavtrtkaedgsndvilmhigfdtnlngthfnplkkgdevkagelicefdidaik 613

QY 421 AAGYEVTTPIVVSNNYKKTGPVNTYGLGIEAGANLLNVAKKEAVPATP 468

DB 614 aagyevttpivvsnnykktgpvntyglgieaganllnvakkeavpatp 661













PR	27-NOV-2000; 2000US-253625P.
PR	22-DEC-2000; 2000US-257931P.
PR	16-FEB-2001; 2001US-269308P.
XX	(ELIT-) ELITRA PHARM INC.
XX	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI	Yamamoto RT, Xu HH;
XX	WPI; 2001-611495/70.
DR	N-PSDB; AAS33075.
XX	New polynucleotides for the identification and development of
PT	antibiotics, comprise sequences of antisense nucleic acids -
XX	Example 3; Seq ID No 10809; 51bp; English.
PS	The invention relates to antisense inhibitors of genes essential to
XX	prokaryotic cellular proliferation, their use in identifying the
CC	genes, their use in the discovery of novel antibiotics, the essential
CC	genes themselves and the encoded proteins. The prokaryotes used are
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC	invention is also useful for the identification of potential new targets
CC	for antibiotic development. The antisense nucleic acids can also be used
CC	to identify proteins used in proliferation, to express these proteins,
CC	and to obtain antibodies capable of binding to the expressed proteins.
CC	The proteins can be used to screen compounds in rational drug discovery
CC	programmes. The antisense nucleic acid sequence is also useful to screen
CC	for homologous nucleic acids which are required for cell proliferation in
CC	a wide variety of organisms. The present sequence represents an
CC	essential prokaryotic cellular proliferation protein.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 484 AA;
SQ	
Query Match	14.3%; Score 338.5; DB 22; Length 484;
Best Local Similarity	33.2%; Pred. No. 6.6e-25;
Matches 94; Conservative 54; Mismatches 114; Indels 21; Gaps 10;	
QY	2 AMVF--PSLVNGYDVAAATMAAG---EMPWMSLEGLDVAAGAGQGTVLPVLVVSWILATI 55
Db	201 asvfatgpal-ggvigavntmlgmppadapnsifgtgtlsag-qggilgfvafawllsl 258
QY	56 EKFLHRLKCTADFLTIPVLTLLTGTFTTAIGPAMRWGVDVLAHGLOGLYDF---GG 111
Db	259 ekqlhkivpesidivtptlslivglatlflmp----vagalsnglvqinvlek99 314
QY	112 PVGGLLFGLVYSPIVTGLHQSPFPIELELFNONGG-SFIFPATASMANIAOGAACLAFFEL 170
Db	315 mvsgftlgltfplmvmgfblhqiltphiemintgmtlllpilamagagvggaalalwir 374
QY	171 AKSEK-LKGLAGASGVAVLGTETPAIFOVNLRWRPFPGITGTRAIGGALTALF-NIK 227
Db	375 cksdkklvemikgalpy-gilgigepliaygtvplgrpfctacigggigavgafgnvg 433
QY	228 AVAAGAAGFLGVVSIDAPDMVMFLVCNAVTTTFATGAAYGL 270
Db	434 aidagsvalipaliannqwlayvigllaayaggfvatiffgi 476
RESULT 12	
AU336679	
ID AU336679 standard; Protein; 484 AA.	
XX AU336679;	
AC AC	
XX	
DT 14-FEB-2002 (first entry)	
XX	

DE Staphylococcus aureus cellular proliferation protein #849.  
XX Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX

OS Staphylococcus aureus.  
XX

PN WO200170955-A2.  
XX

PD 27-SEP-2001.  
XX

PF 21-MAR-2001; 2001WO-US09180.  
XX

PR 21-MAR-2000; 2000US-191078P.  
XX

PR 23-MAY-2000; 2000US-206848P.  
XX

PR 26-MAY-2000; 2000US-207727P.  
XX

PR 23-OCT-2000; 2000US-242578P.  
XX

PR 27-NOV-2000; 2000US-253625P.  
XX

PR 22-DEC-2000; 2000US-257931P.  
XX

PR 16-FEB-2001; 2001US-269308P.  
XX

PA (ELIT-) ELITRA PHARM INC.  
XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX Yamamoto RT, Xu HH;  
XX

DR WPI: 2001-611495/70.  
XX

DR N-PSDB; AAS54538.  
XX

PT New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX

PS Example 3; Seq ID No 12272; 511pp; English.  
XX

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 484 AA;  
SQ

Query Match 13.7%; Score 323; DB 22; Length 484;  
Best Local Similarity 32.2%; Pred. No. 2.4e-23;  
Matches 96; Conservative 45; Mismatches 115; Indels 42; Gaps 10;

QY 9 VNGYDVAATMAAGEMPWLSFLGLDVAQACGCTVLPVIVVSNILATIEFLHRLKGTAD 68

DB 209 v19ggtlllgagknlimvftgeplqpg-qggilgfvavwllsivekrhkhvpaaid 267

QY 69 FLITPVLTLTGFTTGAIGPAMRWGDVLAHGLQGLYDFGPGVGLLGLVYSPVIT 128

DB 268 iivtptiallvglltiffmplagvdsdsvsvngiisigvfgfllgafipvlmi 327

QY 129 GHQSFPPTELEFNO-GGSFTFATASMANIAQGAACLAFFFLAKSEKL--KGLAGASGV 185

DB 328 ghhiiftphiemngsatyilpiaamagagvggaalwrcrnttlrntlkgalpy 387

QY 186 SAVLIGITEPAIFGVNLRWPPFI-----CIGTAAIGCALIALFNKAVAGFLGVV 240  
DB 388 -gflgigepliygtlplgrpflitacigggigavigg----ighigakaipsg-vsl 441  
QY 241 SIDAPDMVMFLVCVVVTFFAFGAAIAYGL---YLVRRNGSIDPDATAAPVPAGTTKA 295  
DB 442 plisdmylygiagll-----aayaggfvcylyf-----gttka 475

RESULT 13

AAU38201  
ID AAU38201 standard; Protein; 455 AA.

XX AAU38201;

DT 14-FEB-2002 (first entry)

DE Salmonella typhi cellular proliferation protein #92.

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Salmonella typhi.

PN WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

DR N-PSDB; AAS56060.

XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX

PT Example 3; Seq ID No 13794; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 455 AA;

Query Match 12.9%; Score 305.5; DB 22; Length 455;  
Best Local Similarity 31.3%; Pred. No. 1.2e-21;  
Matches 79; Conservative 55; Mismatches 111; Indels 7; Gaps 5;

QY 14 VAATMAAGEMPMSLFGDLVQAQYQGVTVLPVLVSVMLATIEFLHKLKGTADFLITP 73  
Db 194 maglispqleqitfg-ealqpg-rggviavilvvalmcwlerqfklipgsllelinp 251

QY 74 VLTLLTGLFTFAIGPAMRWGVDVLAHQGLQGLYDFGPGVGLLFGVSPIVITGLHQS 133  
Db 252 llttvitgavaivalpqlggwdsiahgawaidrggflvgavlagtflplvlihlhgq 311

QY 134 FPPIELELEN-QGGSFTFATASMANIAOGAACLAFFFLAKSEKLKGLAGASGVSAVLGIT 192  
Db 312 lvxlhvelvqahxynalfpillamagvgqigaaiaavlmlktrnarlkvikgalpvgllig 371

QY 193 EPAIFGVNLRWPFIFIGTAAITGGALIALFNKAV-ALGAAGFLGVVSDAPDMWFL 251  
Db 372 eplifgtlpgkpfigacigvggallisykvatvitfgisglplaltivagkvfyl 431

QY 252 ---VCAVVTFPI 260  
Db 432 lgyliaviagfi 443

RESULT 14  
AAU34326  
ID AAU34326 standard; Protein; 687 AA.  
AC AAU34326;  
XX  
XX  
XX 14-FEB-2002 (first entry)  
XX Staphylococcus aureus cellular proliferation protein #602.  
DE  
XX Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX Staphylococcus aureus.  
XX WO200170955-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-207727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX N-PSDB; AAS52185.

XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX Example 3; Seq ID No 5822; 51lpp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 687 AA;

Query Match 12.7%; Score 300; DB 22; Length 687;  
Best Local Similarity 22.0%; Pred. No. 7.9e-21;  
Matches 124; Conservative 88; Mismatches 181; Indels 170; Gaps 21;

QY 17 TMAAGEMPMSLFGDLVQAQYQGVTVLPVLVSVMLATIEFLHKLKGTADFL----IT 72  
Db 127 sqaakepahalvlgptlqgtvgfimgalaaw---cynkfynitlppflgffagkrfv 183

QY 73 PVLTLTGLFTFAIGPAMRWGVDVLAHQGLQGLYDF----GGPVGGLFLGVSPIVIT 128  
Db 184 pvtsvva-----iatgvllsfawppiqdginslnflnklntlttffiglierslipf 238

QY 129 GLHOSF-PPIELEL---FNOGGSFIFATAS--MANIAQG----- 161  
Db 239 glhhifyspfefgfsytnhagelvrqdriwmaqlkdgvftagafgtkypfmmfglp 298

QY 162 AACLAFFFLAKSEKLK---GLAGASGVSANL-GITE-----PAIFGVNLRWLPF 207  
Db 299 aaafayknarperkkvvggimlsagitafitgitteplefsflfvapvlyghvllagts 358

QY 208 FIGIGTAAIGGALIALFNKAVALGAAGFL-----GVVSDAPDMWPF----LVCVVTF 258  
Db 359 fl-----vmblllgvklgmftfsggfidylygllnwdrshallvpgvyaivvy 408

QY 259 FIAGGAIAYGLYLVRRNGSIDPD-----ATAAPVP----- 289  
Db 409 fl-fdfair--kfkltktpredeeteirnsavaklpfdvldamggkenikhldacitr 465

QY 290 -----AG-----TTKAEAE 299  
Db 466 vevvdkskvdvagikaigagvlevgnnmqaifgpkdqikhdmakimsgeltkpsettv 525

QY 300 PAEFSND-----STIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQLVSPV 351  
Db 526 teamsdepvhvealgttdiyapgigqilpisevpdqvfagkmmgdgvgfipekeivapf 585

QY 352 SGKIVVAPPSGHAFVTRTKAEDGNSVDILMHIGFTDTVNLNTHFNPLKKQDEVKAGELL 411  
Db 586 dgtvktlftpkhaigl----esesgvevlhigtdvklngegfeslinvdekvtqgqpl 641

QY 412 CEFDIDAIRKAGYEVTPIVVSN 434  
Db 642 mkvnlaylkhapsivtptmiitn 664

RESULT 15  
AAU37349  
ID AAU37349 standard; Protein; 719 AA.  
XX  
AC AAU37349;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Staphylococcus aureus cellular proliferation protein #1519.  
XX

KW Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.

OS Staphylococcus aureus.

XX WO200170955-A2.

PN

XX 27-SEP-2001.

XX

XX 21-MAR-2001; 2001WO-US09180.

XX

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253825P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

DR N-PSDB; AAS55208.

XX

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

PT

XX

PS Example 3; Seq ID No 12942; 511pp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

QY 17 TMAAGEMPMWSLFGDLVDAQYQGTVPVLVYVSWILATIEKFLKRLKGTADFL-----IT 72

Db 158 sqakeapahalvlgltgtvgfgingalaaw---cynkfynitlpflgffagkrfv 214

QY 73 PVLTLGLLFTFIAGPAMRWGDVLAHGLQGLYDF-----GGPVGGLFGLVYSPVIT 128

Db 215 pivtsvwa-----iatgvllsfawppigdglnslknlttttffiglierslipf 269

QY 129 GLHQSF-PPIELEL---FNQGSFIFATAS--MANIAQG----- 161

Db 270 glhhlfysfwfefsytynahelvgdqrlwmaqlkdgvptagftgkypfmmfglp 329

QY 162 AACLAVFFLAKSEKLLK---GLAGAGVSAYL-GITE-----PAIFGVNLRLRWPF 207

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 330 aaafaiyknarperkkvvgmlsagltflgtiteplefsflfvapvlyghvllagts 389

QY 208 FIGIGTAAIGGALIALFNKAVAGAGFL-----GVSIDAPDMYMF-----LVCAVVTTF 258

Db 390 fl-----vmhllgvkigmtfsggfilyllnwdrshallivpvgivyaivpf 439

QY 259 FTAFGAATAGLYLVRRNGSIDPD-----ATAAPVP----- 289

Db 440 fl-fdfair--xkiktptgredeeteirnsavaklpfdvldamgkknkhldacitrlr 496

QY 290 -----AG-----TTKAEAEA 299

Db 497 vevvdkskvdvagikalgasgvlevgnmqalfigfksdqikhdmakimsgelckpsettv 556

QY 300 PAEFSND-----STIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPV 351

Db 557 teemsdepvhvealgttdiyagqigqilplsevpdgvfagkmmgdvgvfipekgeivapf 616

QY 352 SGKIVVAFPSGHAFVTRKAEODGSNVDIILMHIGFDVNLNGTHFNPLKKGODEVKAGELL 411

Db 617 dgtvktifttkhaigl-----esesgvevllhigldtvkxnggefeslinvdekvttgqpl 672

QY 412 CEFVDIDAIRKAAGYEVTTPIVWSN 434

Db 673 mkvniaylkahapsivtpmiitn 695

Search completed: September 25, 2002, 06:00:08  
Job time: 4034 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:22:04 : Search time 24.86 Seconds  
(without alignments)  
459.822 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MANVFPSPVNGYDVATMAA.....IEAGANLLNVAKEAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/1/1aa/PTCUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113.5	4.8	1176	1	US-07-828-788A-12
2	113.5	4.8	1176	1	US-08-356-034-2
3	113.5	4.8	1176	5	US-08-933-891-2
4	113.5	4.8	1176	5	PCR-US92-1137-12
5	113.5	4.8	1179	1	US-08-040-751-1
6	108.5	4.6	731	2	US-08-911-364-1
7	108.5	4.6	733	4	US-08-464-700-2
8	108	4.6	1011	3	US-08-836-325-2
9	108	4.6	1989	3	US-08-836-325-12
10	107	4.5	1984	3	US-08-836-325-10
11	106.5	4.5	1129	6	5164180-6
12	106.5	4.5	1179	6	5188960-2
13	106.5	4.5	1864	2	US-08-804-227C-3
14	105.5	4.5	462	3	US-08-801-344-9
15	105.5	4.5	462	4	US-09-498-599-9
16	103.5	4.4	396	2	US-08-850-880-4
17	103.5	4.4	396	2	US-08-944-918-4
18	103.5	4.4	396	2	US-08-814-877-4
19	102	4.3	3033	1	US-07-925-695-9
20	101	4.3	3033	1	US-07-925-695-8
21	100.5	4.3	412	4	US-09-445-472-1
22	100.5	4.3	522	4	US-08-894-818B-3
23	100.5	4.3	522	4	US-09-445-472-4
24	100.5	4.3	525	1	US-08-356-340-2
25	100.5	4.3	525	1	US-08-786-555-2
26	100.5	4.3	654	4	US-08-894-818B-35
27	100.5	4.3	654	4	US-09-445-472-16

28 100.5 4.3 713 3 US-09-335-409-11 Sequence 11, Appl  
29 100.5 4.3 713 4 US-09-568-102-11 Sequence 11, Appl  
30 100.5 4.3 713 4 US-09-567-969-11 Sequence 11, Appl  
31 100.5 4.3 713 4 US-09-568-480-11 Sequence 11, Appl  
32 100.5 4.3 713 4 US-09-568-486-11 Sequence 11, Appl  
33 100.5 4.3 713 4 US-09-568-472-11 Sequence 11, Appl  
34 100 4.2 528 2 US-08-403-852D-21 Sequence 21, Appl  
35 100 4.2 528 3 US-08-510-646B-22 Sequence 22, Appl  
36 100 4.2 528 4 US-09-231-818-21 Sequence 21, Appl  
37 100 4.2 1969 3 US-08-836-325-16 Sequence 16, Appl  
38 99.5 4.2 792 2 US-08-678-039A-40 Sequence 40, Appl  
39 99 4.2 756 4 US-08-960-048-10 Sequence 10, Appl  
40 99 4.2 1275 3 US-09-120-513-2 Sequence 2, Appl  
41 98 4.1 1184 6 5254799-6 Patent No. 5254799  
42 98 4.1 1188 6 5254799-7 Patent No. 5254799  
43 97.5 4.1 2005 3 US-08-836-325-7 Sequence 7, Appl  
44 96.5 4.1 593 6 5523211-1 Patent No. 5523211

## ALIGNMENTS

RESULT 1  
US-07-828-788A-12  
; Sequence 12, Application US/07828788A  
; Patent No. 5273746  
; GENERAL INFORMATION:  
; APPLICANT: PAYNE, JEWEL M.  
; APPLICANT: HICKLE, LESLIE A.  
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES  
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID R. SALIWANCHIK  
; STREET: 2421 N.W. 41st STREET, SUITE A-1  
; CITY: GAINESVILLE  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/828,788A  
; FILING DATE: 19920129  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SALIWANCHIK, DAVID R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA75  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1176 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: BACILLUS THURINGIENSIS  
; STRAIN: AIZAWAI  
; INDIVIDUAL ISOLATE: PS811  
; IMMEDIATE SOURCE:  
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK  
; CLONE: 811A2  
; US-07-828-788A-12







```
Db 362 GVGGGTPTGVGAGGPGFPGVGGVGGTIPGVAGVPSVGGVPGVGVPGVGLSPENAAAA 421
QY 288 -----VPAGTTKAEAPAEBSNDSTIIQAPLTGEATIALSSVSDFASG-KLGSGV 338
Db 422 KAAKYGVGTAAAAAARAKAAGFGLVPGVGVAGVPGVGVAGVGLAPGVGAPGV 481
QY 339 AIVPTTGQLVSPVSGKIVVAFPSGHAFVTKAEDGSDNVDILMHIGFDT-----VNL 390
Db 482 GVAPGVG--VAPGIG-----PGVAAAARAKAAKAAQRAAAGLIGAGIPGLGVGVG 533
QY 391 NG-----THFNPLKKGDEVKAGEL-----LCEFD 415
Db 534 PGLGVGAGVGLGVGAGVPGFAGAGDEGVRRLSPELREGDPSSQHLPSTPSPRVPGA 593
QY 416 IDAIIKAAGYEVTPIVVSNTYKTKPNTYGLGEIEAGANLLNVAKEAVPA 466
Db 594 LAAAKAARYGAAVPGVGLGGLGALGVGIPG-GVVGAGPAAAAAARAKA 643

RESULT 7
US-08-464-700-2
; Sequence 2, Application US/08464700
; Patent No. 6232458
; GENERAL INFORMATION:
; APPLICANT: WEISS, ANTHONY S
; APPLICANT: MARTIN, STEPHEN L
; TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,700
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL6520
; FILING DATE: 22-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL9661
; FILING DATE: 28-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00655
; FILING DATE: 16-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GHC3USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-464-700-2
```

Query Match 4.6%; Score 108.5; DB 4; Length 733;  
Best Local Similarity 20.6%; Pred. No. 0.032;  
Matches 122; Conservative 47; Mismatches 227; Indels 195; Gaps 22;

```
QY 16 ATMAAGEMPMSLEGLDVAQAGTGTVLPLVSVWILATIEKFLHKLKGTADFLITPVL 75
Db 110 AGVKPGKVPVGVGLPGV-----YPGVLP-----GARFPVGV----- 140
QY 76 TLLTGLTETIATIGPAMRWYDVLHGLQGLYDFGPGVGLLFLG-LVSPITVITGLHQSF 134
Db 141 --VLPGVPTGANGYKPKAPGVGGAF-A-GIPGVPEGPGQPVLPYPIKAPKLPGGYGLPY 197
QY 135 PPIELELFNOGGSPFIATASMANIAQG-----AACLAVFFLAKSEKILKGLAGA 182
Db 198 TTGKLP-YGYGPGGVAGAGKAGYPTGTGVGPQAAAAAARAKAAGAGVLPVGVGA 256
QY 183 SGYSANLIGITETPAIFGVNLRRLRWFFFIGTAAIGGALIALFNKAKVALGAGEL----- 237
Db 257 -GVPGVPGAI-PGIGGI-----AGVGTAAAAAARAKAAGAGVLPVGGP 303
QY 238 -----GVVSDIDAPDMVMFL----- 251
Db 304 GFGPGVGVPGAGVPGVPGCAGIPVVPVGGAGIPGAAVPGVSPAAAAAARAKAAGYGNRP 363
QY 252 ---VCVVVTFIAPGAAIAYGLYLVRNG-----SIDPDATAAP-- 287
Db 364 GVGVGIPTYGVGAGGPGFPGVGVGGIPGVAGVPSVGGVPGVGVPGVGVGISPEAQAAAA 423
QY 288 -----VPAGTTKAEAPAEBSNDSTIIQAPLTGEATIALSSVSDFASG-KLGSGV 338
Db 424 KAAKYGVGTAAAAAARAKAAGFGLVPGVGVAGVPGVGVAGVGLAPGVGAPGV 483
QY 339 AIVPTKGQLVSPVSGKIVVAFPSGHAFVTKAEDGSDNVDILMHIGFDT-----VNL 390
Db 484 GVAPGVG--VAPGIG-----PGVAAAARAKAAKAAQRAAAGLIGAGIPGLGVGVG 535
QY 391 NG-----THFNPLKKGDEVKAGEL-----LCEFD 415
Db 536 PGLGVGAGVGLGVGAGVPGFAGAGDEGVRRLSPELREGDPSSQHLPSTPSPRVPGA 595
QY 416 IDAIIKAAGYEVTPIVVSNTYKTKPNTYGLGEIEAGANLLNVAKEAVPA 466
Db 596 LAAAKAARYGAAVPGVGLGGLGALGVGIPG-GVVGAGPAAAAAARAKA 645

RESULT 8
US-08-836-325-2
; Sequence 2, Application US/08836325
; Patent No. 6110672
; GENERAL INFORMATION:
; APPLICANT: Mandel, Gail
; APPLICANT: Haleboua, Simon
; APPLICANT: Borden, Laurence A.
; TITLE OF INVENTION: Peripheral Nervous System Specific
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Usin
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave., N. W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,325
; FILING DATE: 2-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: PCT/US95/14251  
FILING DATE: 02-NOV-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/482,401  
FILING DATE: 07-JUN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/334,029  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1011 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-836-325-2

Query Match 4.6%; Score 108; DB 3; Length 1011;  
Best Local Similarity 20.7%; Pred. No. 0.06;  
Matches 91; Conservative 67; Mismatches 168; Indels 114; Gaps 20;

QY 28 LFGLDVAAGYQGTVPVLVSVWILATIEKFLHKLKGTADFL-----ITPVLTLTGT 81  
DB 553 IFDLVTNQA-EDITIMVLICLNMVMTMVEK-----EGQTEYMDVVLHWINNVFIILFTG 605  
QY 82 FLTFTAIQPMRWGCDVLHAGLQGLYDFGPGVGLLFGVSPVITGLHQSFPPIEL 141  
DB 606 ECVLKLIS-----LRH-----YYP--TVG---WNILYFVVVLSIVGMFLAEMIEK 646  
QY 142 FNOGGSFTFATASMANIAQAACLAFFLAKSEKUKLAGAGSVAVIGITEPAIFGVNL 201  
DB 647 Y-----FVSPTLFRVIRLARIGIRL-----IKGAKGIRTLFLALMSPALFNLGL 694  
QY 202 RLWPFPI-----GIGTAA-----IGGALIALFNKAVAGAGPLGVVS 241  
DB 695 LFLVMFIYAFGMSNFAYVVKKEAGINDMNFETFGNSMCLFQITTSAGWGLLAPIL 753  
QY 242 IDAP-----DMVPLVCAVVTFFIAFGAAIAYGLYLVRNGS 278  
DB 754 NSAPDCDPKVKHVGSSVEGDCGNPSVGIFVSVIIISFLVWNMYIA-----VILENFS 809  
QY 279 IDPDATAAPVAGTAKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASGLGV 338  
DB 810 VATEESTEPLESDDEPFEMFVWEKFDPOAT-----QTFEFCCLSD--FAAALDPPLL 859  
QY 339 AIVPTKGOLVS-----PVSGKIVWAPPFSGHAFVTRTKAEDGNSVDIL---MHIGFDTVNLN 391  
DB 860 IAKPNKVLIIANDLPWVSGSDRIHCLDILFATKRVLGE--GGEMDSLRSQMEERFMSANPS 918  
QY 392 GTHFNP-----LKKQGEVKA 407  
DB 919 KVSYPEITTLKRRKQEEVSA 938

RESULT 9  
US-08-836-325-12  
Sequence 12, Application US/08836325  
Patent No. 6110672  
GENERAL INFORMATION:  
APPLICANT: Mandel, Gail  
APPLICANT: Halegoua, Simon  
TITLE OF INVENTION: Peripheral Nervous System Specific  
TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,  
TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational  
TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using

TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C  
STREET: 1100 New York Ave., N. W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,325  
FILING DATE: 2-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14251  
FILING DATE: 02-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/482,401  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/334,029  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0917:0240002  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1989 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-836-325-12

Query Match 4.6%; Score 108; DB 3; Length 1989;  
Best Local Similarity 20.7%; Pred. No. 0.17;  
Matches 91; Conservative 67; Mismatches 168; Indels 114; Gaps 20;

QY 28 LFGLDVAAGYQGTVPVLVSVWILATIEKFLHKLKGTADFL-----ITPVLTLTGT 81  
DB 1508 IFDLVTNQA-EDITIMVLICLNMVMTMVEK-----EGQTEYMDVVLHWINNVFIILFTG 1560  
QY 82 FLTFTAIQPMRWGCDVLHAGLQGLYDFGPGVGLLFGVSPVITGLHQSFPPIEL 141  
DB 1561 ECVLKLIS-----LRH-----YYP--TVG---WNILYFVVVLSIVGMFLAEMIEK 1601  
QY 142 FNOGGSFTFATASMANIAQAACLAFFLAKSEKUKLAGAGSVAVIGITEPAIFGVNL 201  
DB 1602 Y-----FVSPTLFRVIRLARIGIRL-----IKGAKGIRTLFLALMSPALFNLGL 1649  
QY 202 RLWPFPI-----GIGTAA-----IGGALIALFNKAVAGAGPLGVVS 241  
DB 1650 LFLVMFIYAFGMSNFAYVVKKEAGINDMNFETFGNSMCLFQITTSAGWGLLAPIL 1708  
QY 242 IDAP-----DMVPLVCAVVTFFIAFGAAIAYGLYLVRNGS 278  
DB 1709 NSAPDCDPKVKHVGSSVEGDCGNPSVGIFVSVIIISFLVWNMYIA-----VILENFS 1764  
QY 279 IDPDATAAPVAGTAKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASGLGV 338  
DB 1765 VATEESTEPLESDDEPFEMFVWEKFDPOAT-----QTFEFCCLSD--FAAALDPPLL 1814  
QY 339 AIVPTKGOLVS-----PVSGKIVWAPPFSGHAFVTRTKAEDGNSVDIL---MHIGFDTVNLN 391

Db 1815 IAKPNKVLQIADLPWVSGDRIHCLDILFAFTRKRVLGE--GGENDSLRSQMEERFMSNPS 1873  
QY 392 GTHFNP-----LKKQGDEVKA 407  
Db 1874 KVSYPEITTTTLKRKQEVSA 1893

RESULT 10  
US-08-836-325-10  
; Sequence 10, Application US/08836325  
; Patent No. 6110672  
; GENERAL INFORMATION:  
; APPLICANT: Mandel, Gail  
; APPLICANT: Halegoua, Simon  
; APPLICANT: Borden, Laurence A.  
; TITLE OF INVENTION: Peripheral Nervous System Specific  
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,  
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational  
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C  
; STREET: 1100 New York Ave., N. W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836,325  
; FILING DATE: 2-MAY-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/14251  
; FILING DATE: 02-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/482,401  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/334,029  
; FILING DATE: 02-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0917.0240002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1984 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-836-325-10

Query Match 4.5%; Score 107; DB 3; Length 1984;  
Best Local Similarity 20.7%; Pred. No. 0.22;  
Matches 91; Conservative 66; Mismatches 169; Indels 114; Gaps 20;

QY 28 LFGLDVAQAGYQGTVPVLVWVSHILATIEKFLHKLKGTADFL-----ITPVLTLTLTG 81  
Db 1506 IFDLVNTQA--FDITIMVLCLANMVTMVK-----EGQTEYMDYVLHWINMVFILFTG 1558

QY 82 FLTFIAIGPAMRWGDVLAHGLQGLDFGPGVGLLFGVLVSPIVITGLHQSPFPIELEL 141

Db 1559 ECVLKLIS-----LRH-----YVF--TVGWNIFDFVWILSVGMFLA-----EM 1596  
QY 142 FNOGGSFIFATASMANIAQGAACLAFFFLAKSEKLKGLACAGSVASVLGITEPAIFGVNL 201  
Db 1597 IEK-----YFVSPTLFRVIRLARIGRILRL-----IKGAKGIRTLALLFALMMSLPALFNIGL 1647  
QY 202 RLRWPEFI-----GIGTAA-----IGGALIALFNKAKAVALGAAGFLGWVS 241  
Db 1648 LLELVMPYIAFOWSNFAYVYKKEAGINDMFNPFETGNSMCLFQITTTSA--GWGGLLAPIL 1706  
QY 242 IDAP-----DMVMFLVCAVVTFFIAFGAAIAYGLYLVRNGS 278  
Db 1707 NSAPPCDCKKHPGSSVEGDCGNPSVGIYFVSYIIISLVVVMYIA-----VILENFS 1762  
QY 279 IDPDATAAPVPACTTAAEAAPAEFSDNSTIIQAPLTGEAIALSSVSDAMFASGKLGSV 338  
Db 1763 VATEESTEPLESDDFEMFVEVWEKFDPDAT-----QFIECKLSD--FAAALDPPLL 1812  
QY 339 AIYPTKGQVLS---PVSGRIVVAFPSGHAFVARTKAEDGSNVDIL---MHIGFDTVTNLS 391  
Db 1813 IAKPNKVLQIADLPWVSGDRIHCLDILFAFTRKRVLGE--GGENDSLRSQMEERFMSNPS 1871  
QY 392 GTHFNP-----LKKQGDEVKA 407  
Db 1872 KVSYPEITTTTLKRKQEVSA 1893

RESULT 11  
5164180-6  
; Patent No. 5164180  
; APPLICANT: Payne, Jewel-Sick, August J.  
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE  
; AGAINST LEPIDOPTERAN PESTS  
; NUMBER OF SEQUENCES: 6  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/451,389  
; FILING DATE: 14-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 353,860  
; FILING DATE: 18-MAY-1989  
; SEQ ID NO: 6  
; LENGTH: 1129  
5164180-6

Query Match 4.5%; Score 106.5; DB 6; Length 1129;  
Best Local Similarity 22.4%; Pred. No. 0.1;  
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 249 MFLVCAVVTFFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 294  
Db 380 LFVLDTGTEFSASLTADLPSTIY--RQGTVDSDLVIPQDQNSVPARAGFSHRLSHVTML 437  
QY 295 AEA-----EAP-----AEFSN--DSTIIQAPLTGEATLSSVSDAMFASGKLGS 336  
Db 438 SQAAGAVYTLRAPTFWSHRSAEFLNIPSSQITQIPLT--KSNLMSGTSVYKGPGETGG 496  
QY 337 GVAIVPTKGQVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGFDTVTNLTGTFN 396  
Db 497 DILRRTPSGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINOCNFS 551  
QY 397 PLKKQGEVKGAGELLCEFFDIDAKAAGYEVTTTIVSVNKKTKGPVNTYLGETEAGANLL 456  
Db 552 ATMSSGNGNLQSG-----SFRTAGF--TTPF---NFSNGSSIFTLARHFNVNSG-NEV 596  
QY 457 NVAKKEAVPA 466  
Db 597 YIDRIEFVPA 606

RESULT 12  
5188960-2  
; Patent No. 5188960







Search completed: September 25, 2002, 06:00:58  
Job time: 2334 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:55:04 : Search time 45.29 Seconds  
(without alignments)  
992.930 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVPSLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1058.5	44.8	651	2 S44257	phosphotransferase
2	953.5	40.4	664	1 B32243	phosphotransferase
3	759.5	32.1	627	2 H95951	fusion, PTS system
4	701	29.7	480	2 S39978	scrA protein - Sta
5	695.5	29.4	630	2 H83686	PTS system, beta-g
6	695	29.4	617	2 AB1167	phosphotransferase
7	674	28.5	633	2 AC1436	PTS system, beta-g
8	673.5	28.5	636	2 C83724	PTS system, beta-g
9	673	28.5	632	2 S68599	phosphotransferase
10	673	28.5	655	2 C95220	trehalose PTS syst
11	667	28.2	634	2 AD1078	PTS system, beta-g
12	664	28.1	627	2 F95200	PTS system IABC c
13	660	27.9	627	2 E98067	phosphotransferase
14	659	27.9	479	2 F82432	PTS system, sucros
15	659	27.9	480	2 D90038	PTS system, sucros
16	658	27.8	705	2 A99084	phosphotransferase
17	657	27.8	609	2 I40406	beta-glucoside per
18	646	27.3	479	2 J00781	sucrose uptake pro
19	645	27.3	609	2 T47097	hypothetical prote
20	640	27.1	631	2 B42603	beta-glucoside-spe
21	628	26.6	617	2 AC1421	beta-glucoside-spe
22	612	25.9	628	2 D97073	PTS system, beta-g
23	588.5	24.9	636	2 D86807	hypothetical prote
24	580	24.5	618	2 AC1204	phosphotransferase
25	578.5	24.1	612	2 B97935	hypothetical prote
26	570.5	24.1	612	2 C69725	hypothetical prote
27	570	24.1	470	2 C69725	phosphotransferase
28	569.5	24.1	470	2 H83926	PTS system, trehal
29	567.5	24.0	625	2 C25977	phosphotransferase

30	509	21.5	640	2 AB1423	beta-glucoside-spe
31	496	21.0	475	2 C89813	hypothetical prote
32	489	20.7	456	2 S62331	phosphotransferase
33	487	20.6	458	2 H83881	PTS system, sucros
34	473	20.0	455	1 WQEBST	phosphotransferase
35	467	19.8	372	2 I39868	sac operon regulat
36	468	17.3	494	2 AG1231	PTS system trehalo
37	406	17.2	459	2 JU0293	levansucrase synth
38	396.5	16.8	494	2 AF1585	PTS system trehalo
39	390	16.5	460	2 A39938	phosphotransferase
40	367	15.5	473	2 C65236	phosphotransferase
41	367	15.5	473	2 A98281	trehalose specific
42	367	15.5	483	2 AI0449	protein-Npi-phosph
43	360	15.2	473	2 A86122	PTS system enzyme
44	348.5	14.7	681	2 A89781	hypothetical prote
45	345	14.6	665	2 B96970	PTS enzyme II, ABC

ALIGNMENTS

RESULT 1

S44257  
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus pen  
C:Species: Pediococcus pentosaceus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 16-Jul-1999  
C:Accession: S44257  
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.  
submitted to the EMBL Data Library, April 1994  
A:Description: The sucrose and raffinose operons of Pediococcus pentosaceus pPE1.0.  
A:Reference number: S44252  
A:Accession: S44257  
A:Molecule type: DNA  
A:Residues: 1-651 <LEE>  
A:Cross-references: EMBL:Z32771; NID:g493728; PIDN:CAA83668.1; PID:g475968  
C:Genetics:  
A:Gene: scrA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
C:Keywords: phosphotransferase  
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 44.8%; Score 1058.5; DB 2; Length 651;  
Best Local Similarity 46.7%; Pred. No. 1.2e-65;  
Matches 221; Conservative 78; Mismatches 151; Indels 23; Gaps 5;

QY 1 MAMVPSLVNGYDVAAATMAAGEMPMSLFGDVAQAGYQGTVPVLVYVSWILATIERFLH 60  
Db 194 MINVLPVLNGYSVAITMAAGKMYNVNVLGHLVAQAGYQGVLPVLGVAFILATLERFFH 253  
QY 61 KRUKGTADFLITPVLVLLLTGFLTFTTAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFLG 120  
Db 254 KHKGAFDEFTTFMFAIVITGFTFTIVGFLRTVSDALTNGLVGLYNTSGWGMGIFGL 313  
QY 121 VYSPIVTGLHQSPPTTELELF---NQGSFIFATASMANIAQGAACLAFFLAKSEKL 176  
Db 314 LYSIVITGLHQTFPALETLLANVAKTGSGFIFPVASMANIGOGAATLAIFATRSQKQ 373  
QY 177 KGLAGASGVSALGITEPAIFGVNLRWRPFIFIGTAAIGGALIALFNKAKVALGAAGF 236  
Db 374 KALTSSAGVSALLGITEPAIFGVNLRMKKPPFVAATASGASAFGLFLHVLVSVAMGPASV 433  
QY 237 LGVVSIDAPDMWFLVCVVVTFIAFGAAIAYGLYLVRNGSIDPDATAAPVPAGTTKAE 296  
Db 434 IGFISIAKSIAPAFMLSAVISFVFAFTPTFIY----AKRTLGDDRQVKSPATSTV--- 486  
QY 297 AEPAPFSDNSTIIQAPLTGEATLSVSDAMFASGKLSGVAIVPTKGLVSPVSKIV 356  
Db 487 -----INVNDEIISAPVTGASESLKQVNDQVFSAEITMGKGAIVPSSDQVWAPADGVIT 540  
QY 357 VAPPSGHAFVTRKAEODGSNVVDILMHIGFDVNLNGHFNPLKKQGVKAGELLCEFDI 416  
Db 541 VTDSHHAYGIKTTA----GAELIHLGLDVTNLNGEHEFTTNVQKGDVTHQGDLTGTFDI 596



Matches 146; Conservative 49; Mismatches 81; Indels 8; Gaps 4;

QY 1 MAMVPSLVNGYDVAATMAAG-EMPWWSLFGLDVAQAGYQGTVPVLVSVSWILATIEKFL 59  
Db 194 MILVHPBELMSAYDYPKALEAKREIPHNWLFGEINQVGQGVPLMLVAFIILATIEKGL 253  
QY 60 HRLKGTADFLIPPVLTLLLTGTLFTIAIGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFG 119  
Db 254 RKVIPVLDNLLTPLLAILSTGTFITSEFVGPLTFTLGWLSDGLTWLYERGGAGIGLIFG 313  
QY 120 LYSPIVITGLHQSPPPIELF-----NOGGSFIFATASMANIAQGAACLAFFLAK-SE 174  
Db 314 LLVAPIVITGMHSHFAIETQLIADSSSTGGSFIFPIATMSNIAQGAALAAFFLIKENK 373  
QY 175 KKLKLAGASVSAVLGITEPAIFGVNLRWRPFIFIGTAAIGGALIALFNKAVAGLAA 234  
Db 374 KLGVASAGVSALLGITPEPMFGVNLKURPIFGIAGVSGIGSAYIAFFKVAIALGTA 433  
QY 235 GFLGVYSIDAPD--MVNMLFCAVVTFIAFGAAIAYGLYLVRN 276  
Db 434 GIGFISISQNGWLHYGIAIIAIFVAGVTYALSRYKKYEN 477  
RESULT 5  
H83686  
PTS system, beta-glucoside-specific enzyme II, ABC component BH0296 [imported] - Bacillus  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: H83686  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: H83686  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-630 <STO>  
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04015.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: H83686  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 29.4%; Score 695.5; DB 2; Length 630;  
Best Local Similarity 32.8%; Pred. No. 1.5e-40;  
Matches 152; Conservative 108; Mismatches 182; Indels 21; Gaps 5;

QY 2 AMVFPSSLVNGYDVAATMAAGEMPWWSLFGLDVAQAGYQGTVPVLVSVSWILATIEKFLHK 61  
Db 185 ALIYPSIVELHDSRIDV-----TFEGIPVVMNYTSTVFPIILAVFAMSVFEKCNK 236  
QY 62 RLKGTADFLITPVLTLTLTGFLFTIAIGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFGVL 121  
Db 237 KIHEAKNFTLTPMCLMLIVPLTFLAAGPLGTFTISQGLASGYTFIYNLSPIVAGAFMGAF 296  
QY 122 YSPITVITGLHQSPPPIEL-ELFNOGGSFIFATASMANIAQGAACLAFFLAKSEKLGLA 180  
Db 237 KIHEAKNFTLTPMCLMLIVPLTFLAAGPLGTFTISQGLASGYTFIYNLSPIVAGAFMGAF 296  
QY 122 YSPITVITGLHQSPPPIEL-ELFNOGGSFIFATASMANIAQGAACLAFFLAKSEKLGLA 180  
Db 297 WQVLVIFGIHWGIVPTIMNLSRYGRDTMIAMVGPNSFAQAGASLGVLTKTKPEVKAIA 356  
QY 181 GASVSVAVLGITEPAIFGVNLRWRPFIFIGTAAIGGALIALFNKAVAGLAAFL--- 237  
Db 357 GSAALTGFGGITPEPSIYGVTLKYKKPEVIAIAGIAGVAGGSSGAANAIPGILTLP 416  
QY 238 -----GVVSDAPDMVNFCAVVTFIAFGAAIAYGLYLVRNNGSIDPATAAPVPAGT 292  
Db 417 IFIGKGFVGIILGAVAYIILSAIGTYFFGYKDEMGDI-----APT 457  
QY 293 TKAEAPAEFSNDSTIIQAPLTGEATLSVSDAMFASGKLGSGVAIVPTKGOLVSPVS 352  
Db 458 TKAEKGTVE---AEVIVSPIRGNIVPLNEVKDEAFSAGLLGKGAIVPQEGKLISPVN 513  
QY 353 KGIWVAFPSGHAFAVRTKAEDGSNVDILMHIGFTVNLNGTHFNPLKKGDEVKAGELLC 412  
Db 514 GTIETAFPTGHAIGIRS-----DRGVEILLHVGFDTVOLNGKYFKLLVLAQGRDLVVGQALL 569  
QY 413 EFDIDAIIKAAGYEVTTPTPIVVSN 434  
Db 570 EFDLEAIKADGYDITPTPIVVTN 591  
RESULT 7  
AC1436  
PTS system, beta-glucosides specific enzyme IIABC homolog lin0026 [imported] - Lister

Db 532 LFETKHAIGITS-----DNGTEIFIHVGIDIVQLKGEHFTSFIEQGDVAAGDVLLFEDE 587  
QY 418 AIRAAGYEVVTPIVSNYKKTGPNVTYGLGEIEAGANLLNAK 460  
Db 588 RITAAGYDVITPVLITNAKQFSNVQTTDKREVTSEDLLIHVK 630

RESULT 6  
AB1167  
phosphotransferase system (PTS) beta-glucoside-specific enzyme IIABC component homolo  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AB1167  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1167  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-617 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98816.1; PID:g16410127; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0738  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 29.4%; Score 695; DB 2; Length 617;  
Best Local Similarity 34.8%; Pred. No. 1.6e-40;  
Matches 154; Conservative 82; Mismatches 162; Indels 44; Gaps 6;

QY 2 AMVFPSSLVNGYDVAATMAAGEMPWWSLFGLDVAQAGYQGTVPVLVSVSWILATIEKFLHK 61  
Db 185 ALVYPTMINLFNEGAHITFLQIP-----VVLMSYSFSFVPIILAVNLSILERFLNS 236  
QY 62 RLKGTADFLITPVLTLTLTGFLFTIAIGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFGVL 121  
Db 237 KIHEAKNFTLTPMCLMLIVPLTFLAAGPLGTFTISQGLASGYTFIYNLSPIVAGAFMGAF 296  
QY 122 YSPITVITGLHQSPPPIEL-ELFNOGGSFIFATASMANIAQGAACLAFFLAKSEKLGLA 180  
Db 297 WQVLVIFGIHWGIVPTIMNLSRYGRDTMIAMVGPNSFAQAGASLGVLTKTKPEVKAIA 356  
QY 181 GASVSVAVLGITEPAIFGVNLRWRPFIFIGTAAIGGALIALFNKAVAGLAAFL--- 237  
Db 357 GSAALTGFGGITPEPSIYGVTLKYKKPEVIAIAGIAGVAGGSSGAANAIPGILTLP 416  
QY 238 -----GVVSDAPDMVNFCAVVTFIAFGAAIAYGLYLVRNNGSIDPATAAPVPAGT 292  
Db 417 IFIGKGFVGIILGAVAYIILSAIGTYFFGYKDEMGDI-----APT 457  
QY 293 TKAEAPAEFSNDSTIIQAPLTGEATLSVSDAMFASGKLGSGVAIVPTKGOLVSPVS 352  
Db 458 TKAEKGTVE---AEVIVSPIRGNIVPLNEVKDEAFSAGLLGKGAIVPQEGKLISPVN 513  
QY 353 KGIWVAFPSGHAFAVRTKAEDGSNVDILMHIGFTVNLNGTHFNPLKKGDEVKAGELLC 412  
Db 514 GTIETAFPTGHAIGIRS-----DRGVEILLHVGFDTVOLNGKYFKLLVLAQGRDLVVGQALL 569  
QY 413 EFDIDAIIKAAGYEVTTPTPIVVSN 434  
Db 570 EFDLEAIKADGYDITPTPIVVTN 591

RESULT 7  
AC1436  
PTS system, beta-glucosides specific enzyme IIABC homolog lin0026 [imported] - Lister

C:Species: *Listeria innocua*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AC1436  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1436  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-633 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC95259.1; PID:g16412447; GSPDB:GN00178  
A:Experimental source: strain Clil1262  
C:Genetics:  
A:Gene: lin0026  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 28.5%; Score 674; DB 2; Length 633;  
Best Local Similarity 36.2%; Pred. No. 4.6e-39;  
Matches 161; Conservative 75; Mismatches 179; Indels 30; Gaps 10;

Qy 2 AMVPSL--VNGYDVAATMAAG---EMPMSLF-GLDVAQAGYQGTVPVLVWSMILATI 55  
Db 181 ALVYPTLAGISAGDPYTLFAGTIFESPIHVTFLGIPILMSYASSVIPILATVFGSKV 240  
Qy 56 EKFLHKLKRGADFLITPVLILLTGFTTFAIGPAMRWGVDVLAHGLQGLVDFGPPVGG 115  
Db 241 EKGEKKIIPDVIKTVPVPCFLLVVPVITFLVIGPIATWAGQLLGGATWVYNLSPVAVG 300  
Qy 116 LFLGLVSPVITGLHQSPFPIEL-ELFNQGGSFIFATASMANIAQGAACLAFFLAKSE 174  
Db 301 LILGFWQVFIPLHGLIPVAINLLTGLSDPVLAMFAGSPAQICAVILAVFFTRNK 360  
Qy 175 KLKLAGASGVSALVIGTEPAIFGNLRLRPFIFIGITAAIGGALIALFNKIKAVLAGAA 234  
Db 361 KIKLSIPAFISGIGVTEPAIYGTTLPLKPFIMSCVAGAVGGIIGFVSAKVIYMGSL 420  
Qy 235 GLGVVSDAP-----DMVFLVCAVTVFFTAFCAGAAIYGLYLVRNGSIDPDATAAPVP 289  
Db 421 GIGLPNFPFGAGITSFAFWVTAIVTSFI-----LGFILAYVAGFKDP---AEAVV 470  
Qy 290 AGTTKAEAPAEAFSNDSTIIQAPLTGEALSSVDAMFASGLGSGVAIVPTKGOLVS 349  
Db 471 BETNTEGETLIE----RETIPAPVVGELVLDVKDAFSSGALGKGVAILPTVGRVVA 526  
Qy 350 PVSGKIVVAFPSGHAFVTRKAEDGSNVDIILMHIGFDVNLNGTHFNPLKQGDVKAGE 409  
Db 527 PAAGVTVTTFPTGHAIGITT--NDGA--EVLHIGMDTVQLGKFFTAHVKGQDVIEKQ 582  
Qy 410 LLCEDDIAIKAAGYEVTPIVSN 434  
Db 583 LTFEDIEGKAAGYDVTTPVVTN 607

RESULT 8  
C83724  
PTS system, beta-glucoside-specific enzyme II, ABC component bglp [imported] - *Bacillus*  
C:Species: *Bacillus halodurans*  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: C83724  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: C83724  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-636 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04314.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: bglp  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 28.5%; Score 673.5; DB 2; Length 636;  
Best Local Similarity 34.0%; Pred. No. 5e-39;  
Matches 163; Conservative 78; Mismatches 193; Indels 45; Gaps 9;

Qy 2 AMVPSLVNGYDVAATMAAGEMPMSLF-----GLDVAQAGYQGTVPVLV 47  
Db 183 SLVYPTLV-----VLTEGE-PLYTLEGTIFESPVHITFLGIPILMSYATSVIPIIL 234  
Qy 48 VSWILATIEKELHKLKAGTADFLITPVLTLTLLTGFTTFAIGPAMRWGVDVLAHGLQGLY 107  
Db 235 AAYFASKVEARLKIIPDVVKTFLPVFFTLVIVLPLTFLVIGPIATWAGQLLQGFLLVY 294  
Qy 108 DFGPGVGLLFLGLVYSPVITGLHQSPFPIEL-ELFNQGGSFIFATASMANIAQGAACLA 166  
Db 295 NLSPIIAGAFGLGGFWQVFVIFGLHWGLPIAINNLLVVOGSDPVLAMVFAASFAQIGAVAA 354  
Qy 167 VFFLAKEKLKLAGASGVSALVIGTEPAIFGNLRLRPFIFIGITAAIGGALIALFNI 226  
Db 355 VMLKIKQOKVKTLSVPAPFISGIFGVTPEPAIYGVTLPLKRPFIISCIAAAVGAAGIIGLFRS 414  
Qy 227 KAVLAGAGELGVWSI----DAPDMVME-LVCVVVTFIAGCAATAYGLYLVRNGSIDP 281  
Db 415 QGYIIGLIGIFGIFSLHPADGMDAGFWGVIATVAVAV- LGFILTLYLFLGKSGNASDEQ 473  
Qy 282 DATAAPVAGTTKAEAPAEAFSNDSTIIQAPLTGEALSSVDAMFASGLGSGVAIV 341  
Db 474 TETRAHTSTGTGKEE-----ISSPFNGSVITLSEIKDEAFSSGALGSGIAIE 521  
Qy 342 PTKGOLVSPGKIVVAFPSGHAFVTRKAEDGSNVDIILMHIGFDVNLNGTHFNPLKQ 401  
Db 522 PSEGLFSPGVMYALYPTTHALGITT----DRGAELLIHGLDVTQLDGKFFTAHTIQ 577  
Qy 402 GDEVKAGLLECFDIDAIAKAAGYEVTPIVSNYKTKTPVNTYGLGTEAGANLLNVAK 460  
Db 578 GAQVEKGLLIEFDIKEIKAAGYAVTTPVIVNHKQYQLFLTDKQVNAGRDLLELTR 636

RESULT 9  
S68599  
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - *Streptococcus*  
N:Alternate names: sucrose-specific enzyme II  
C:Species: *Streptococcus sobrinus*  
A:Variety: strain 6715  
C>Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 07-May-1999  
C:Accession: S68599  
R:Chen, Y.Y.M.; Lee, L.N.; LeBlanc, D.J.  
Infect. Immun. 61, 2602-2610, 1993  
A:Title: Sequence analysis of *scrA* and *scrB* from *Streptococcus sobrinus* 6715.  
A:Reference number: S68598; MUID:93273516  
A:Accession: S68599  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-632 <CHE>  
A:Cross-references: EMBL:L06791  
C:Genetics:  
A:Gene: *scrA*  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
C:Keywords: phosphotransferase; sugar transport system  
F:480-632/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 28.5%; Score 673; DB 2; Length 632;  
Best Local Similarity 36.4%; Pred. No. 5.4e-39;  
Matches 168; Conservative 76; Mismatches 197; Indels 20; Gaps 8;

Qy 1 MAMVPSLVNGYDVAATMAAGEMPMSLFGLDVAQAGYQGTVPVLVYVSWILATIEKFLH 60

Db 191 LMVNPALPNAYAV---ASDAKALTFGFG-IPVVGYGQGTVLPAFPVGMIGARLENWLH 245  
QY 61 KRLKGTADFLITPVLTLLLTGFTTFAIGAMPARWVGDLAHGLOGLYDFGPGVGGLLFGL 120  
Db 246 KRYPEALDLITLFTFLVNSILGLFAIGPVFHSVETVLAATAEWILALPFGIAGIIGG 305  
QY 121 VYSPIVITLHQSFPPTELELEFGOGSFI-ATASMANIAOGAACLAVFLAKSEKLKGL 179  
Db 306 LOQVIVVTGVHHLNFLETLQALAEKAPENPNLLSATAAGVAVLAVAKTSAKUKAL 365  
QY 180 AGAGSVAVLGTTEPAIFGVNLRWPFIFIGTGAATGALIALFNKIKALVAAGFLGV 239  
Db 366 AYPALSAAIGITEPAIFGVNLRWPFVGMVGGSGAGFIALVGLKATGMSVTVLPGL 425  
QY 240 VSDADPWFLVCVVTFIATGAATAYGLYLVRNGSIDDPATAPVPAGTTKABAEA 299  
Db 426 LLFLNSQPMYIIVSITVACAI-ALT-YYFGYADKEEDVSAKKPEAPAAVAETET 481  
QY 300 PAEFSNDSTIIQAPLTGEATALSVDAMFASGLGSGVAIVPTKGLVSPVSGKIIVAF 359  
Db 482 KSE-----VIASPLDGEAVELSKVNDPVFSSBAMGKGIAPKPSGNTVYSPVNGTVQIAF 535  
QY 360 PSGHAFAVRTKAEDGNSVDILMHIGEDTVNLNGTHFNPLKQKQDEVKAGELLCEFDIDAI 419  
Db 536 ETGHAYGLA--KSDNGA--EVLHVGIDTVSMNGTGFQDKVAANQTVKVGDLGTFSDAKI 591  
QY 420 KAAGYEVTTPVIVSNYKTKGPVNTYGLGEIEAGANLLNVAK 460  
Db 592 AEAGLDOTTVIIINTADYSEVRPLAAGLAHAGLDLLELNK 632  
RESULT 10  
C95220  
trehalose PTS system, IIABC components [imported] - Streptococcus pneumoniae (strain TIGR:SP4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: C95220  
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Elsen, J.A.; Read, T.D.; Peterson, S.; Helton, J.D.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: C95220  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-655 <KUR>  
A:Cross-references: GB:AE005672; PIDN:ANK75956.1; PID:g14973388; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: Spt1884  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 28.5%; Score 673; DB 2; Length 655;  
Best Local Similarity 35.3%; Pred. No. 5.7e-39;  
Matches 170; Conservative 87; Mismatches 173; Indels 52; Gaps 16;  
QY 1 MAMVPSLVNGYDVAATMAAGNP--WMSLFGLDVAQAGYQGTVPVLVLYVSWILATIEKF 58  
Db 195 ICIVSQLLNAYAVASPPAADIANNWNNPGYFTVNRIGYQAOVIPALLAGLSLYLEIF 254  
QY 59 LHKRLAGTADFLITPVLTLLLTGFTTFAIGAMPARWVGDLAHGLOGLYDFG--GPVG-- 114  
Db 255 WHKHIEPVISMIVFPLSLIPALIAHTVLGP----IGWTIGGGLSVVLAGLTGPVKWL 310  
QY 115 -GLLFGLVYSPVITLHQSFPPTELELEFGOGSFI-ATASMANIAOGAACLAVFLAK 172  
Db 311 FGAIFGALYAPVITGLHHTMTNIDTQLIADAGGTALWPMIALSNIAQSGAVFAYFMHR 370  
QY 173 -SEKLAGAGAGSVAVLGTTEPAIFGVNLRWPFIFIGTGAATGALIALFNKIKAL 231

Db 371 HDREAQVSLPAISAYLGVTEPALFGVNVKYYIPFVAGMTGSALACMLSVTFNVTAASI 430  
QY 232 GAAGFLGVVSDIDAPDMVMF----LVCAVVTFFIAFGAAIAYGLYLVRNGSIDDPDATAAP 287  
Db 431 GIGLPGILSIQPOQYMLPFAAGTMLVAIVPMLLTF-----FFRKAGLFTK----- 475  
QY 288 VPAGTTKAEAE----APAEFSNDSTI-----TQAPLTGEAIALSSVSDAMFASGLKLSG 337  
Db 476 -TEGDNLQAEFVAQAEAEFVNHEPVELTSVEIISPLTGQVKELSOATDPIFASGVNGQG 534  
QY 338 VAIVPTKGLVSPVSGKIIVAFPSGHAFVARTKAEDGNSVDILMHIGEDTVNLNGTHFNP 397  
Db 535 LVTEPSOGELTSPNGIVTVLPFKHAIGI--VSDEG--VELLIHIGMDTVGLDGKGFES 590  
QY 398 LKQKQDEVKAGELLCEFDIDAIKAAGYEVTTPVIVSN---YKKTGPVNTYGLGEIEAGAN 454  
Db 591 LVVQGDHVTGQQILRFDMVIRKAGLVTTETPVIITNQDAYTATIP-GTYPT-TIQAGAS 648  
QY 455 LL 456  
Db 649 LM 650  
RESULT 11  
AD1078  
PTS system, beta-glucosides specific enzyme IIABC homolog lmo0027 [imported] - Lister  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AD1078  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1078  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-634 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98242.1; PID:g16409386; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0027  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 28.2%; Score 667; DB 2; Length 634;  
Best Local Similarity 34.6%; Pred. No. 1.4e-38;  
Matches 156; Conservative 81; Mismatches 172; Indels 42; Gaps 10;  
QY 2 AMVFPSSL--VNGYDVAATMAAG--EMPMSLFLGLDVAQAGYQGTVPVLVLYVSWILATI 55  
Db 182 ALVITLAGITAGDPLVTLFAGTIFESPPIHVTFLGIPVILMSYASSVPIIIVATYFGSV 241  
QY 56 EKFLHKLKAGTADFLITPVLTLLLTGFTTFAIGAMPARWVGDLAHGLOGLYDFGPGVGG 115  
Db 242 EKGFKKIIPDVITKTVFPVFCILLIVPITFIVGPIATWAGQLLGAGTIWYNLSPIAG 301  
QY 116 LLFGLVYSPVITLHQSFPPTELELEFGOGSFI-ATASMANIAOGAACLAVFLAKSE 174  
Db 302 LILGGFQWQVFIQGLHWGLVPVAINLTLVGHDPILANTWFCASFAQIGAVLAVFFKSRNK 361  
QY 175 KKLGLAGAGSVAVLGTTEPAIFGVNLRWPFIFIGTGAATGALIALFNKIKALVAAG 234  
Db 362 KIKLSIPAFISGIFGVTEPAIYGTPLPKPFIMSGIAGIGGIIIGFAGSQTIIYMG 421  
QY 235 GFLGVVYSIDAP-----DMVMFLVCVVTFIATGAATAYGL-----YLVRNGSIDPDA 283  
Db 422 GIGLPLNFFPGSGISGEFFWVVIIVISFILTYVVGFKDPADVVEQSNVTE--- 478

QY 284 TAAPVAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIYPT 343  
Db 479 -----GETLIERET-----IPAPVVGIEVTLADVKDEAFSGALGKGVAIIP 521  
QY 344 KGQLVSPVSGKIVVAFPSCHAFVAKTRAEKDSNVDLMHIGFDTVNLGTHFNPLKKQGD 403  
Db 522 VGRVAPAGVTTIIPFGHAIGITTK--DGA--EVLHIGMDTVQLEKGKFTTAHVKQGD 577  
QY 404 EVKAGELCEFDIDAIAKAAGYEVTTPVIVSN 434  
Db 578 VIEKQQLTEFDIEGKAAGYDVTTPVVVTN 608  
RESULT 12  
F95200  
PTS system IIBC components [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: F95200  
R:Tectelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95200  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75799.1; PID:g14973217; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 28.1%; Score 664; DB 2; Length 627;  
Best Local Similarity 34.3%; Pred. No. 2.3e-38;  
Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;  
QY 1 MAMVPPSLVNGYDYAATNAAGEMPWLSLFLGDLVAQAGYQGVTLPLVLYVSWILATIERFLH 60  
Db 182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGQLGSLVFAFIIGVVGAKFEKAVR 237  
QY 61 KRLKGTADFLITPVLTLTLTGLFTIAIGPAMRWGDVLAHGLQGLYDFGPGVGLLFLG 120  
Db 238 KVPDPVIDLTVTPFVLLVMSILGLFVIGPVHVVENVILIATRAILSMPTGLGGLIGG 297  
QY 121 VYSPVITGLHQSPPPIELELFNQGGSFIF-ATASMANIAQACLAIVFFLAKSEKLG 179  
Db 298 VHQLIVSGVHHIFNLLEVLQLAADHANPFAITTAAMTAQGAATVAVGVKTKNPKLKL 357  
QY 180 AGASGVSAVLGITTEPAIFGVNLRWRPFFIGTAAIGGALIALFNKAKVALGAAGFLG- 238  
Db 358 APPAALSFLGITTEPAIFGVNLRWRPFFKPFSLIAGIAGGLASILGLAGTNGTIIIPGT 417  
QY 239 VVSDAPDMWMLVCAVVTFFIAFCAATAYGLYLVRNGSIDPDATAAPVPAGTTKAEAE 298  
Db 418 MLYVNGQLPOYLLMVAVSFALGFALTYMFGY-----EDEVATAAKAEVAEKEE 470  
QY 299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIYPTKGQLVSPVSGKIV 357  
Db 471 VAPAAQLNETLV--TPIVGDVVALADVNDPVFSSGAMGGIIVKVPSCGVVYAPADAESVI 528  
QY 358 APPSGHAFVAKTRAEKDSNVDLMHIGFDTVNLGTHFNPLKKQGDVEVKAGELCEFDID 417  
Db 529 APFTGHAFGLKTR----NGAEVLHVGIDTVSMNGDGFETKVAQGNKVKAGDVLGTFDSN 584  
QY 418 AKAAGYEVTTPVIVSN---YKKTGPVNTYGLGEIEAGANLLN 458  
Db 585 KIAAAGLDDTTWIVTNGDYASVAPVAT---GSVAKGDVAVIEV 625



A:Reference number: A82035; MUID:20406833

A:Accession: F82432

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <HEI>

A:Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0653

A:Map position: 2

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match

Best Local Similarity 27.9%; Score 659; DB 2; Length 479;

Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

QY 1 MAMVFPVSLVNGYDVAATMAAGEPMWLSFLGDLVAQAGYQGTLPVLVYVSWILATIEKFLH 60  
 DB 194 MLMVHPDLLNGMGFGSASVGTPTNILGFEIEKVGQGVLPVLVSAYILAKIENGLR 253  
 QY 61 KRUKGTADFLITPVLTLTGLTFTTAIGPAMRWGDVLAHGLQGLYDFGGPVGGLLFLG 120  
 DB 254 KIVPSVVDNLLTPMLAIFITFTVVGPLTRDVGFMGLDNLNWLDSAGFVGGALFGF 313  
 QY 121 VYSPVITGLHOSFPPIELELF---NOGGSFIFATASMANIAOAGACLAVFFLAKSEKL 176  
 DB 314 ITAPFVITGMHSHFIAIEIQLADIYTTGTFIFPFAASNTAQAALAVGVMTRETKL 373  
 QY 177 KGLAGASVSALVIGITEPAIFGNLRWRPFIFIGITAAIGGALIALFNKIKAVAGAAAGF 236  
 DB 374 KGVAIFSGVTALLGITEPAMFGVNLKLRYPFIAAICGAALASAFITLNVKAQALGAAGL 433  
 QY 237 LGVWSIDAPDMVNFVLCVVTFIAPGAAL 266  
 DB 434 PGIISINPOQIGYIMGMAISFVAAPALTIV 463

RESULT 15

D90038

PTS system, sucrose-specific IIBC component [imported] - Staphylococcus aureus (strain N

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: D90038

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ul, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11416146

A:Accession: D90038

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-480 <KUR>

A:Cross-references: GB:BA000018; PID:g13702328; PIDN:BAB43469.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: Scra

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match

Best Local Similarity 27.9%; Score 659; DB 2; Length 480;

Matches 139; Conservative 53; Mismatches 82; Indels 12; Gaps 5;

QY 1 MAMVFPVSLVNGYDVAATMAAGE-PMWLSFLGDLVAQAGYQGTLPVLVYVSWILATIEKFL 59  
 DB 194 MILVHPSLSAYDFPKAVENGKAIPTVDVGLHINQVGGQVLPMLVAYILASIEKGL 253  
 QY 60 HRRKGTADFLITPVLTLTGLTFTTAIGPAMRWGDVLAHGLQGLYDFGGPVGGLLFG 119  
 DB 254 RKVIPVLDNLLTPLLSIFITAFITLTSFVGPIITRLQGLWLSDGLTWLYEFGGAIGGLIFG 313

QY 120 LVYSPVITGLHOSFPPIELELF---NOGGSFIFATASMANIAOAGACLAVFFLAK-SE 174  
 DB 314 LLYAPVITGMHSHFIAVETTLIADATKTCGSPFIPIATMSNVQAAGAAIAAFIHKONK 373  
 QY 175 KKLGLAGASVSALVIGITEPAIFGNLRWRPFIFIGITAAIGGALIALFNKIKAVAGAA 234  
 DB 374 KKLGVASAAAGISALLGITEPAMFGVNLKLRYPFIGAIVGSGIGSAVIAPFKVKAIALGTA 433  
 QY 235 GFLGVVSIDA--PDMVNFVLCVVTFIAPGAALVGLYLVRNGS 278  
 DB 434 GLPGFISINPVAHAGLHYFVGMTISFII-----AITVTLLILSKRKN 475

Search completed: September 25, 2002, 06:02:01

Job time: 417 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2002, 06:01:04 ; Search time 25.03 Seconds  
(without alignments)  
723.961 Million cell updates/sec

Title: US-09-604-231-2

Perfect score: 2363

Sequence: 1 MAMVPSLVNGYDVATMAA.....IEAGALLNVAKKAVPATP 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059.5	44.8	651	1 PTSA_PEDPE	P43470 pediococcus
2	953.5	40.4	664	1 PTSA_STRMU	P12655 streptococ
3	701	29.7	480	1 PTSEB_STAXY	P51184 staphylococ
4	678	28.7	474	1 PTSEB_PASMU	O9c122 pasteurella
5	657	27.8	609	1 PTBA_BACSU	P40739 bacillus su
6	646	27.3	479	1 PTSEB_VIBAL	P22825 vibrio algi
7	640	27.1	631	1 PTBA_ERWCH	P26207 erwiniia chr
8	570	24.1	470	1 PTBA_BACSU	P39794 bacillus su
9	567.5	24.0	625	1 PTBA_ECOLI	P08722 escherichia
10	492	20.8	456	1 PTSEB_SALTY	P08470 salmonella
11	489	20.7	456	1 PTSEB_KLEPN	P27219 klebsiella
12	462.5	19.6	674	1 PTGA_CORGL	Q45298 corynebacte
13	406	17.2	459	1 SACK_BACSU	P15400 bacillus su
14	390	16.5	460	1 PTSEB_BACSU	P05306 bacillus su
15	367	15.5	473	1 PTTB_ECOLI	P36672 escherichia
16	321.5	13.6	699	1 PTGA_BACSU	P20166 bacillus su
17	317.5	13.4	651	1 PTAA_KLEPN	P45604 klebsiella
18	299	12.7	726	1 PTGA_STRPN	P35595 streptococ
19	298.5	12.6	648	1 PTAA_ECOLI	P09323 escherichia
20	294.5	12.5	324	1 PTGA_BACST	P42015 bacillus st
21	287.5	12.2	631	1 YBFS_BACSU	P39816 bacillus su
22	270.5	11.4	485	1 PTDA_ECOLI	P24241 escherichia
23	256	10.8	189	1 PTGA_BORBU	Q44840 borrelia bu
24	255	10.8	168	1 YPOE_BACSU	P50829 bacillus su
25	254.5	10.8	634	1 LACY_STRTR	P23936 streptococ
26	250.5	10.6	474	1 YFEV_ECOLI	P77272 escherichia
27	249.5	10.6	154	1 PTGA_MYCCA	P45618 mycoplasma
28	246	10.4	168	1 PTGA_ECOLI	P08837 escherichia
29	244.5	10.3	482	1 PTSEB_VIBCH	Q9kvd9 vibrio chol
30	244	10.3	168	1 PTGA_SALTY	P02908 salmonella
31	238.5	10.1	165	1 LACY_HAEIN	P45338 haemophilus
32	233	9.9	627	1 LACY_LACDE	P22733 lactobacill
33	231.5	9.8	161	1 PTGA_BUCAL	Q9wx17 buchnera ap

34	230.5	9.8	940	1 PTGA_MYCPN	P75569 mycoplasma
35	221	9.4	641	1 RAPP_PEDPE	P43466 pediococcus
36	204	8.6	908	1 PTGA_MYCGE	P47315 mycoplasma
37	165.5	7.0	639	1 LACY_LEULA	Q48624 leuconostoc
38	159	6.7	483	1 PTVB_ECOLI	P32154 escherichia
39	151.5	6.4	658	1 HRSA_ECOLI	P54745 escherichia
40	144.5	6.1	527	1 PTIB_BACSU	P54715 bacillus su
41	136.5	5.8	359	1 PTWC_ECOLI	P36672 escherichia
42	130	5.5	578	1 PTFB_RHOCA	P23387 rhodobacter
43	126.5	5.4	580	1 PTFB_XANCP	P23355 xanthomonas
44	123.5	5.2	632	1 PTMA_BUCAL	P57635 buchnera ap
45	122	5.2	694	1 PTFA_MYCPN	P75039 mycoplasma

#### ALIGNMENTS

```
RESULT 1
PTSA_PEDPE
ID PTSA_PEDPE STANDARD; PRT; 651 AA.
AC P43470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS system, sucrose-specific IIABC component (EIIABC-SCR) (Sucrose-
DE permease IIABC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Pediococcus pentosaceus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Pediococcus
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PPe1.0;
RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXX-1994) to the EMBL/GenBank/DDSI databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: Z32771; CAA83668.1; -
EMBL: L32093; AAR25567.1; -
HSSP: P20166; IGPR.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
```





CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE006222; AAK03930.1; -.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; FALSE\_NEG.  
 DR Phosphotransferase system; Transferase; Sugar transport; Plasmid;  
 KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.  
 FT DOMAIN 1  
 FT EIIB DOMAIN.  
 FT EIIIC DOMAIN.  
 FT TRANSMEM 109 474  
 FT TRANSMEM 109 129  
 FT TRANSMEM 158 178  
 FT TRANSMEM 182 202  
 FT TRANSMEM 229 249  
 FT TRANSMEM 264 284  
 FT TRANSMEM 303 323  
 FT TRANSMEM 345 365  
 FT TRANSMEM 376 396  
 FT TRANSMEM 403 423  
 FT TRANSMEM 444 464  
 FT MOD\_RES 26 26  
 FT MOD\_RES 324 324  
 FT SEQUENCE 474 AA; 49908 MW; D26FA1A059603C01 CRC64;

Query Match 28.7%; Score 678; DB 1; Length 474;  
 Best Local Similarity 49.5%; Pred. No. 1.le-37;  
 Matches 135; Conservative 51; Mismatches 83; Indels 4; Gaps 1;

QY 1 MAMVPSLVNGYDVAATMAAGEMPWNSLGLDVAQAQYQGTIVLPVIVSVWILATIEFLH 60  
 DB 194 MLLVHPALSDGWNVALTLAKGNIQWIFGLIEIRVYQGTIVPVLVSWWVLAFLERKL 253  
 QY 61 KRLKGTADFLLTVLLTGLTFIAIGPMRWGVGLAHGLQGLYDFGPGVGLLFL 120  
 DB 254 KVPVSFLDNLITPLFAFLVGLLAFTVIGPIGRAGSLISTGLTWLDTLGVGGAIFGT 313  
 QY 121 VYSPITVIGLHQSPPIELELFNQ-----GGSFIPATASMANIAGAACLAFLVFLAKSEKL 176  
 DB 314 LYAPIVITGMHQTETIAVETQLLAEVARTGTFIPPIAAMSNIAGAACLAFLVFLAKSEKL 373  
 QY 177 KGLAGASGVSAVLGITPAIFGVNLRWRPFFIGTGTAAICGALLIALFNKAVAGAGP 236  
 DB 374 RGIAPVSGISALLGITPAMPFGVNLRYRYPFISAMIGAGISSAVIALFNKAVIALGAAGL 433  
 QY 237 LGVVSIDAPDMVFLVCAVVTFFTAAGAAIAYG 269  
 DB 434 PGIPSIRPDSLAWVCVGMGLISASIAFTLTVILG 466

RESULT 5

PTBA\_BACSU  
 ID PTBA\_BACSU STANDARD; PRT; 609 AA.  
 AC P40739; Q45661.  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE PTS system, beta-glucoside-specific IABC component (EIIABC-BGL)  
 DE (Beta-glucoside-permease IABC component) (Phosphotransferase  
 DE enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).  
 GN BGLP OR N17C.  
 OS Bacillus subtilis.  
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RN SEQUENCE FROM N.A.

RC STRAIN-168 / MARBURG;  
 RX MEDLINE-95189730; PubMed=7883710;  
 RA Le Coq D., Lindner C., Krueger S., Steinmetz M., Stuelke J.;  
 RT "New beta-glucoside (bgl) genes in *Bacillus subtilis*; the bglP gene  
 RT product has both transport and regulatory functions similar to those  
 RT of BglF, its *Escherichia coli* homolog.";  
 RL J. Bacteriol. 177:1527-1535(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-168 / BGSC1A1;  
 RC MEDLINE-95219088; PubMed=7704263;  
 RX Yoshida K.-i., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;  
 RA "Cloning and sequencing of a 29 kb region of the *Bacillus subtilis*  
 RT genome containing the hut and wapa loci.";  
 RL Microbiology 141:337-343(1995).  
 RN [3]  
 RN SEQUENCE OF 1-182 FROM N.A.  
 RP STRAIN-168;  
 RC MEDLINE-96204517; PubMed=8628237;  
 RX Beloin C., Hirschbein L., le Hegarat F.;  
 RA "Suppression of the Bgl+ phenotype of a delta hns strain of  
 RT *Escherichia coli* by a *Bacillus subtilis* antiterminator binding  
 RT site.";  
 RL Mol. Gen. Genet. 250:761-766(1996).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPP); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 CC histidine + sugar phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z34526; CAA84286.1; -.  
 DR EMBL; D31856; BAA06652.1; -.  
 DR EMBL; D29985; BAA06256.1; -.  
 DR EMBL; X85408; CAA59697.1; -.  
 DR EMBL; Z99124; CAB15963.1; -.  
 DR PIR; S47174; S47174.  
 DR HSP; P20166; IGPR.  
 DR Subtilisin; BGL0934; bglP.  
 DR InterPro; IPR001127; PTS\_EIIB.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIB\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIB; 1.  
 DR PROSITE; PS00371; PTS\_EIIB\_1; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Complete proteome.  
 FT DOMAIN 1 41  
 FT EIIB DOMAIN.  
 FT EIIIC DOMAIN.  
 FT EIIA DOMAIN.  
 FT MOD\_RES 25 25  
 FT MOD\_RES 302 302  
 FT MOD\_RES 532 532  
 FT MOD\_RES 112 112  
 FT TRANSMEM

FT	TRANSEM	141	161	POTENTIAL.
FT	TRANSEM	174	194	POTENTIAL.
FT	TRANSEM	202	222	POTENTIAL.
FT	TRANSEM	246	266	POTENTIAL.
FT	TRANSEM	281	301	POTENTIAL.
FT	TRANSEM	321	341	POTENTIAL.
FT	TRANSEM	351	371	POTENTIAL.
FT	TRANSEM	379	399	POTENTIAL.
FT	TRANSEM	412	432	POTENTIAL.
FT	CONFLICT	75	75	A -> S (IN REF. 3).
FT	CONFLICT	288	288	L -> F (IN REF. 2).
FT	CONFLICT	434	434	E -> G (IN REF. 2).
FT	CONFLICT	436	436	A -> S (IN REF. 2).
FT	CONFLICT	449	450	DG -> HR (IN REF. 2).
FT	CONFLICT	549	549	I -> M (IN REF. 2).
FT	CONFLICT	552	552	G -> S (IN REF. 2).
SQ	SEQUENCE	609 AA;	64550 MW;	5F630C671D21FBED CRC64;

Query Match 27.8%; Score 657; DB 1; Length 609;  
Best Local Similarity 33.4%; Pred.No. 3.4e-36;  
Matches 153; Conservative 84; Mismatches 181; Indels 40; Gaps

QY	13	DVAATWAACEMPWLSFLGLDVAAQAGYQCCTVLPVLVVYSWILATIEKFLHRLKGTADFLIT	72
DB	186	DLTALLGAGR-P-I-SFIGLPPVTAAATSYSSVIPILLSINIATSVASYEKRWDIFRTHASLKLIIV	243
QY	73	PVLTLLTLTGPTAIGAPMRWGDVLAHGLOGLYDFGGPVGGLLFGLVSPVITVGLHQ	132
DB	244	PTFTLLIIVPLTLITVGPLCAILGEVLSGSVNYLFDHAGLIVAMILLAGTFSLIIWTGMHY	303
QY	133	SFPPIELELFINOGG-SFIPATASMANIAQAACLAIVFFLAKSEKLKLAGAGSVAVLGI	191
DB	304	AFVPIMINNIAQNQHGYLLPAMELANMGQAGASFVFLRSRNKKPKSLALTTISITALMGI	363
QY	192	TEPAICGVNLRWRPFPGICTGAACGALIALFNRIKAVALGA-AGFLGVVSDADPDWMVF	250
DB	364	TEPMAYGVNMRLKPPFAALIGGAAGGAFYGMTGVASIYVGGNAGLPSIPVFIPTFIYA	423
QY	251	LVCADVTFETAFGAATAYGGLYVRRNGSIDPDATAAPVPAGTTKAEAEAPAEFSNDSTII	310
DB	424	MIGLVTAFAAETAAYALLGFEDVPSDGS-----QOPAVHEGSRREII	464
QY	311	OAPLTGEATALSVDAMFASGKLGSGVAIVPTKGOLVSPVSKIVVAPPSPGHAFVRTK	370
DB	465	HSPIKEVKALSEVKDGVFSAGVMGKGFAPPEEGEWSPVRGVSVTTFIKTKHAIGITS-	523
QY	371	AEDGSNVDTLMHIGFDTVNLNGHTFNPLPKQGDDEVKAGELCECFIDAIKAAGYEVTTP	430
DB	524	---DOGAETLIHTGLDVTVLEQGWFTHAIKEGDKVAPGDPFLVSFDLEQIKAAGXDVITVP	580
QY	431	VSNKYK--TGPVNTYTGLGIEAGANLLNVAKKEAVPA	466
DB	581	IVTNTQYSFSPVKEIG-----KVQPKEALLA	607

RESULT 6

ID	PTSB_VIBAL	STANDARD;	PRT;	479 AA.
AC	R22825;			
DC	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-			
DE	permease IIBC component) (Phosphotransferase enzyme II, BC component)			
DE	{BC 2.7.1.69} (EIIBC-SCR).			
GN	SCRA.			
OS	Vibrio alginolyticus.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=663;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91071601; PubMed=2174811;			

```

Blatch G.L., Scholle R.R., Woods D.R.;
"Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
uptake-encoding region.";
Gene 95:17-23(1990).
[2]
SEQUENCE OF 1-12 FROM N.A.
MEDLINE-91285433; PubMed-2060795;
Blatch G.L., Woods D.R.;
"Nucleotide sequence and analysis of the Vibrio alginolyticus scr
repressor-encoding gene (scrR).";
Gene 101:45-50(1991).
-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
THE SUGAR.
-1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
histidine + sugar phosphate.
-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
-1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
-1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M76768; AAA27555.1; "-
EMBL; M35009; AAA27557.2; "-
PIR; JQ0781; JQ0781.
HSP; P05053; IIBA.
InterPro; IPR001996; PTS_EIIB.
InterPro; IPR003352; PTS_EIIC.
Pfam; PF00367; PTS_EIIB; 1.
Pfam; PF02378; PTS_EIIC; 1.
ProDom; PD001476; PTS_EIIB; 1.
PROSITE; PS01035; PTS_EIIB_CYS; 1.
Phosphotransferase system; Sugar transport; Transferase.
Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? EIIB DOMAIN.
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 479 AA; 49890 MW; FBF906B5170E3EB7 CRC64;

Query Match 27.3%; Score 646; DB 1; Length 479;
Best Local Similarity 46.8%; Pred. No. 1.4e-35;
Matches 128; Conservative 54; Mismatches 87; Indels 6; Gaps 3;

QY 1 MAMVPSLVNGYDVAATMAAGEMPMSLFLGDVAAQAGYQGTVLPVLVVSWTLATIEKFLH 60
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 MLMVHPDLLNGWGGGASVSGNIPTWNLLEFGIEQKVGQGSVLPVLSAFILAKVELGLR 253
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 KRLKGTADELFTLVLTLLLTGFTFFIATIGPAMRWGVDLAHGLQGLYDFGPGVGLLFLGL 120
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 KVPISVLNLLTLPILAIFAGLLTFTVVGPTTRDGLGLDGLNWLNTAGVGVGAVFGL 313
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 VYSPIVITGLHQSPPTPELELF----NQGSFIFATASMANIAGAACLAFFFLAKSEKL 176
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 IYAPFVITGMHHSFIATETQLADIAITGTGTFIFPIAAMSNSVSGAALAAVGNMSDKKM 373
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 KGLAGAGSVSAVLGITPAILGVNLRLRWPFPIGIGTAAIGGALIALFNKAVALGAAGF 236
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 KGIAIPSGVTLGLGITPAMGVNKLRIYPIFAVCAALASASBITFMNKAQALGAAGL 433
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 LGVVISIDAPMVMFLVCA-VVTFIFAFGAATAYGL 270
||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 434 PGIISI-TPDKIGYIAGMVIATLTAFTVITLGI 467

## RESULT 7

PTBA\_ERWCH  
ID PTBA\_ERWCH STANDARD; PRT; 631 AA.

AC P26207;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE PTS system, beta-glucoside-specific IIABC component (EIIABC-BGL)

DE (Beta-glucoside-permease IIABC component) (Phosphotransferase

DE enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).

GN ARBF.

OS Erwinia chrysanthemi.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI\_TaxID=556;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9211114; PubMed=1732212;

RA el Hassouni M., Henrissat B., Chippaux M., Barras F.;

RT "Nucleotide sequences of the arb genes, which control beta-glucoside

utilization in Erwinia chrysanthemi: comparison with the Escherichia

coli bgl operon and evidence for a new beta-glycosidase family

including enzymes from eubacteria, archaeobacteria, and humans.";

RL J. Bacteriol. 174:765-777(1992).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

CC -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY

CC SIMILARITY)

CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein

CC histidine + sugar phosphate.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

FT TRANSMEM 99 119 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 146 166 POTENTIAL.  
FT TRANSMEM 175 195 POTENTIAL.  
FT TRANSMEM 206 226 POTENTIAL.  
FT TRANSMEM 248 268 POTENTIAL.  
FT TRANSMEM 272 292 POTENTIAL.  
FT TRANSMEM 295 315 POTENTIAL.  
FT TRANSMEM 328 348 POTENTIAL.  
FT TRANSMEM 358 378 POTENTIAL.  
FT TRANSMEM 385 405 POTENTIAL.  
FT TRANSMEM 407 427 POTENTIAL.  
FT TRANSMEM 434 454 POTENTIAL.  
SQ SEQUENCE 631 AA; 66984 MW; 7D0BD27A36BFFDF CRC64;

Query Match 27.1%; Score 640; DB 1; Length 631;  
Best Local Similarity 33.8%; Pred. No. 4.6e-35;  
Matches 154; Conservative 86; Mismatches 200; Indels 16; Gaps 8;

Qy 3 MVFSLVNGYDVAATMAAGEMPMSLSGLDVAQAGYQCTVLPVLVSVILATIEKFLHXR 62  
Db 184 LVHFSMAAFN---AMQAPDHSHTLHFLGIPITFINYSVPIILFASVWVCKLEKPLNRW 240  
Qy 63 LKGTADFLITPVLLTLLTFTFAIGPAMRWGDVLAHGLQGLYDFGPGVGLLGLVY 122  
Db 241 LHANIRNFTPLLCIVISVPLTFLLLIGPSATWLSOMLAGYQWLYGLNLSLAGAVMGALW 300  
Qy 123 SPIVITGLHQSPPELELFNQ-GGSFIFATASMANIAQGAACLAFFLAKSEKLGLAG 181  
Db 301 QVCVIFGLHWGFVPLMLNFSVIGHDTLLPLLPVAVLGOAGATLGLVLLRTQDLKRGKGIAG 360  
Qy 182 ASGVSAVLGITPAIFGVNLRWPFPGTGTAAIGGALIALFNKIALVALGAAGFLGVVS 241  
Db 361 SAFSNAIFGITPAVGVYVTLPLRPFPGTGTGCGALGAAGVAYHTMTYSFGPSPISFTQ 420  
Qy 242 IDAPDMNVLV-CAVVTFFIAG-AAIAYGLYLVRRNGSIDPDATAAPVPAAGTTKAEAE 299  
Db 421 VIPPTGVDSSVMAAVIGTLLAFALTSWFGVPKD---ETQPAADSPA--VLAETQA 475  
Qy 300 PAEFSNDSTIIQPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQVLSPVSGKIVVAF 359  
Db 476 NAGAVRDETLE-SPLAGEVLLLEQVADRTFASGVNGKGIARPTQGRLYAPVDGTVASLF 534  
Qy 360 PSGHAFVARTKAEDGSNNVDILMHIGFDVNLNGTHFNPLKQGVKAGELLCDFDIDAI 419  
Db 535 KTHHAIGLASR---GGAEVLIHVIGIDTVRLDGRYFPHVRVGVVROGDLLEDFDGP 590  
Qy 420 KAAGYEVTTPIVNSYKTKTGPVNTYGLGEIEAGANL 455  
Db 591 EAAGYDLTTPITVITNSEDYRGVPEPVASGVKVDANAPL 626

## RESULT 8

PTTB\_BACSU

ID PTTB\_BACSU STANDARD; PRT; 470 AA.

AC P39794; O34771;

DT 01-FEB-1995 (Rel. 31, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable PTS system, trehalose-specific IIBC component (EIIBC-TRE)

DE (Trehalose-permease IIBC component) (Phosphotransferase enzyme II, BC

DE component) (EC 2.7.1.69) (EII-TRE).

GN TREP OR TREP.

OS Bacillus subtilis.

OC Bacteria, Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / MARBURG;

RX MEDLINE=97074649; PubMed=8917076;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;

RA Schoeck F., Dahl M.K.;



RT \*Analysis of DNA flanking the treA gene of Bacillus subtilis reveals  
 RT genes encoding a putative specific enzyme iitre and a potential  
 RT regulator of the trehalose operon.";  
 RL Gene 175:59-63(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AC327;  
 RX MEDLINE-97124190; PubMed-8969503;  
 RA Yamamoto H., Uchiyama S., Sekiguchi J.;  
 RT Cloning and sequencing of a 4.6 kb segment in the 73 degrees-76  
 RT degrees region of the Bacillus subtilis chromosome containing genes  
 RT for trehalose metabolism and acetoin utilization.";  
 RL Microbiology 142:3057-3065(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AC327;  
 RX MEDLINE-97417488; PubMed-9272861;  
 RA Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;  
 RT Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region  
 RT of the Bacillus subtilis genome reveal genes for a new two-component  
 RT system, three spore germination proteins, an iron uptake system and a  
 RT general stress response protein.";  
 RL Gene 194:191-199(1997).  
 RN [4]  
 RP SEQUENCE OF 324-470 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE-95379486; PubMed-7651129;  
 RA Helfert C., Gotsche S., Dahl M.K.;  
 RT "Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by  
 RT a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";  
 RL Mol. Microbiol. 16:111-120(1995).  
 CC - FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC - CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar -> protein  
 CC histidine + sugar phosphate.  
 CC - SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC - SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
 CC - SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 254245; CAA91014.1; -  
 DR EMBL; D83967; BAA23409.1; -  
 DR EMBL; D86417; BAA22289.1; -  
 DR EMBL; X80203; CAA56494.1; -  
 DR EMBL; 299108; CAB12609.1; -  
 DR HSSP; P05053; 1IBA.  
 DR Subtilist; BG11009; treP.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.  
 FT DOMAIN 1 ? EIIB DOMAIN.  
 FT DOMAIN ? 470 EIIC DOMAIN.  
 FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 110 130 POTENTIAL.  
 FT TRANSMEM 160 180 POTENTIAL.  
 FT TRANSMEM 183 203 POTENTIAL.

FT TRANSMEM 234 254 POTENTIAL.  
 FT TRANSMEM 263 283 POTENTIAL.  
 FT TRANSMEM 301 321 POTENTIAL.  
 FT TRANSMEM 326 346 POTENTIAL.  
 FT TRANSMEM 347 367 POTENTIAL.  
 FT TRANSMEM 375 395 POTENTIAL.  
 FT TRANSMEM 403 423 POTENTIAL.  
 FT TRANSMEM 443 463 POTENTIAL.  
 FT CONFLICT 140 140 F -> S (IN REF. 1).  
 FT CONFLICT 363 363 M -> L (IN REF. 1 AND 4).  
 FT CONFLICT 465 465 A -> G (IN REF. 1 AND 4).  
 SQ SEQUENCE 470 AA; 49999 MW; 7A741850A3697D53 CRC64;  
 Query Match 24.1%; Score 570; DB 1; Length 470;  
 Best Local Similarity 42.8%; Pred. No. 1.4e-30;  
 Matches 116; Conservative 54; Mismatches 99; Indels 2; Gaps 1;  
 QY 3 MYFPSLVNGYDVAAATWAAGEMPMSLFLGDVAQAGQGTPLVPLVYVSWILATIEKFLHKKR 62  
 DB 197 LVHPDLLNANGGAARQSGEIPWNFLGLEVKVGQGVLPILLASYMLAKIEVELTKR 256  
 QY 63 LKGTADFLITPVLTLLLTGTLFTAIATGPMRWVGDYLAHGLQGLYDFGGPVGGLLFLVY 122  
 DB 257 TPEGIQLLVVAPITLLTGTFSPIIGTITFAIGNVLTSGLSIVFGSFAALGGLLYGGFY 316  
 QY 123 SPVITGLHOSPPPIELEFNQ--GGSFIFATASMANIAQGAACLAFLVFLAKSEKUKGLA 180  
 DB 317 SALVITGMHTFLAVDLQLIGSKLGTFLWPMIALSNIAQGSAAALAMMFTVKDEKQKGLS 376  
 QY 181 GASGVSAVLGITPEATFGVNLRLRWPFPGTIGTAATGGALIALFNKAVNALGAAGFLGVV 240  
 DB 377 LSGISAYLIGTPEALFGVNLRYRFFITAWSSGLAGMYISSQGVLLASVGVGGVPGF 436  
 QY 241 SIDAPDMVMLCAVVVTFITFAAGAAIAYGLY 271  
 DB 437 SIMSQYGAFAIGMAIVLIVPFAGTYAYARF 467  
 RESULT 9  
 ID PTBA\_ECOLI STANDARD; PRT; 625 AA.  
 AC P08722.  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE PTS system, beta-glucoside-specific IIABC component (EIIABC-BGL)  
 DE (Beta-glucoside-permease IIABC component) (Phosphotransferase  
 DE enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).  
 GN BGLF OR BGLC OR BGLS OR B3722.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-88009877; PubMed-3309161;  
 RA Bramley H.F., Kornberg H.L.;  
 RT "Nucleotide sequence of bglc, the gene specifying enzymeIIBgl of the  
 RT PEP:sugar phosphotransferase system in Escherichia coli K12, and  
 RT overexpression of the gene product.";  
 RL J. Gen. Microbiol. 133:563-573(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-87222180; PubMed-3034860;  
 RA Schnetz K., Toloczky C., Rak B.;  
 RT "Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide  
 RT sequence, genetic organization, and possible evolutionary  
 RT relationship to regulatory components of two Bacillus subtilis  
 RT genes.";  
 RL J. Bacteriol. 169:2579-2590(1987).

[3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=93315143; PubMed=7686882;  
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;  
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli  
 RL genome: organizational symmetry around the origin of replication.";  
 Genomics 16:551-561(1993).  
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HRP); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON BGLG.  
 CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar - protein  
 CC histidine + sugar phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M15746; AAA83837.1; -;  
 DR EMBL: M16487; AAA23510.1; -;  
 DR EMBL: L10328; AAA62073.1; -;  
 DR EMBL: AE000449; AAC76745.1; -;  
 DR PIR: C25977; C25977.  
 DR PIR: A47616; A47616.  
 DR HSP: P20166; IGPR.  
 DR EcoGene; EG10115; bglF.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR ProSITE; PS00371; PTS\_EIIA\_1; 1.  
 DR ProSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Inner membrane; Complete proteome.  
 FT DOMAIN 1 39 EIIB DOMAIN.  
 FT DOMAIN ? ?  
 FT DOMAIN 495 625 EIIA DOMAIN.  
 FT MOD\_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 306 306 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 547 547 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 100 120 POTENTIAL.  
 FT TRANSMEM 141 161 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 203 223 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 285 305 POTENTIAL.  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 355 375 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT TRANSMEM 404 424 POTENTIAL.  
 FT TRANSMEM 432 452 POTENTIAL.  
 FT TRANSMEM 490 510 POTENTIAL.  
 FT TRANSMEM 526 546 POTENTIAL.  
 SQ SEQUENCE 625 AA; 66482 MW; 9612A2207125C4E6 CRC64;

Query Match 24.0%; Score 567.5; DB 1; Length 625;  
 Best Local Similarity 30.8%; Pred. No. 2.7e-30;  
 Matches 142; Conservative 91; Mismatches 211; Indels 17; Gaps 7;  
 QY 2 AMVPEPSLVNGYDVAAATMAAGEMPWMSLEGLDYAAQAGTQVLPVLVYVSWILATTEKFLHK 61  
 DB 180 ALVHPLILTAFTAPGOKADALGL--DFLIGIPVTLNLYSSSVIPIIFSALWCSILERRLNA 236  
 QY 62 RLKGTADFLIIPVLTLLTGLTFTIAIGPAMRWGVDVLAHGLQGLYDFGGPVGGLLFLV 121  
 DB 237 WLPRAIKNFFPLCLMVTPTVTELLVGPLSTWISLIIAGYLMWYQAVPAFAGVWGPF 296  
 QY 122 YSPITVLGHQSPPIELFN-QGGSFIFATASMANIAQAACLAFFFLAKSEKLKGLA 180  
 DB 297 WQIFVMEGLHWGLVPLCLINNFVTLGYDTMPLMPATMAQVGAALGVFLCERDAQKQVVA 356  
 QY 181 GASVSVALGITEPAIFGVNLRWRPFVIGTAAIGALIALFNKIKAVAGAGFLGV 240  
 DB 357 GSAALTSFLGITEPAVYGVNLPKYPFVIACISGALGATIIGYATQKVYSGFLPSIFTFM 416  
 QY 241 SIDAPDMVFLCAVVTFFIAFGAAIAVGLVLRNGSIDPD-ATAAPVPAGTTKAEAA 299  
 DB 417 QTIPSTGIDFTVWASV-----IGGVIAIGCAFV---GTVMHFITAKRQAGAPQKPTP 468  
 QY 300 PAEFSNDSTIIQAPLTGEATLSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIYVAF 359  
 DB 469 EVITPPQGGICSPMTGEIVPLIHVADTTFASGLLGKGIATLPSVGEVRSVPAGRIASLF 528  
 QY 360 PSGHAFVRVKAEDGSNDVILMHIGFDVTNUNGTHFNPLKQGDENVKAGELLCFEDDAI 419  
 DB 529 ATLHAIGI--ESDDG--VEILIHVGIDITVKLDGKFFSAHVNVGDKVNTGDRLLISFDIPAI 584  
 QY 420 KAAGYEVTTPVWSNKKTPGVNTYGLGEIAGANLLNVAK 460  
 DB 585 REAGFDLTPVLISNSDDFTDVLPHGTAQISAGEPLLSIIR 625  
 RESULT 10  
 ID PTSB\_SALTY STANDARD; PRT; 456 AA.  
 AC P08470;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-  
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)  
 DE (EC 2.7.1.69) (EIIB-SCR).  
 GN SCRA.  
 OS Salmonella typhimurium.  
 OG Plasmid pUR400.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OC NCBI\_TaxId=602;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RN MEDLINE=88216186; PubMed=3285123;  
 RX Ebner R., Lengeler J.W.;  
 RA "DNA sequence of the gene scrA encoding the sucrose transport protein  
 RT EnzymeII (Scr) of the phosphotransferase system from enteric bacteria;  
 RT homology of the EnzymeII (Scr) and EnzymeII (Bgl) proteins.";  
 RL Mol. Microbiol. 2:9-17(1988).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=94018607; PubMed=8412665;  
 RA Jahreis K., Lengeler J.W.;  
 RT "Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid  
 RT repressor for sucrose and D-fructose specific regulons from enteric  
 RL bacteria.";  
 RL Mol. Microbiol. 9:195-209(1993).  
 RN [3]  
 RP SEQUENCE OF 1-7 FROM N.A.  
 RC STRAIN=6153-62;

```

RX MEDLINE=91100329; PubMed=1846143;
RA Hardesty C., Ferran C., Drenth J.M.;
RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT characterization of scry, the structural gene for a
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT outer membrane porin.";
RL J. Bacteriol. 173:449-456(1991).
CC -I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC COVALENTLY BOUND IIA DOMAIN. INSTEAD, EII-SCR-MEDIATED
CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
CC GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC -I- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar - protein
CC histidine + sugar phosphate.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -I- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -I- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67750; CAA47973.1; -
DR EMBL; Y00541; CAA68605.1; ALT_SEQ.
DR EMBL; M38416; AAA98418.1; -
DR PIR; S01036; WOEBST.
DR HSP; P05033; IIBA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
DR Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? EIIB DOMAIN.
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
FT SEQUENCE 456 AA; 47864 MW; 98A6F1620AE50885 CRC64;
CC -----
Query Match 20.8%; Score 492; DB 1; Length 456;
Best Local Similarity 38.5%; Pred. No. 1.8e-25;
Matches 104; Conservative 52; Mismatches 100; Indels 14; Gaps 3;

QY 3 MVFPSLVNGVDVAA---TAAAGEMPMSLFGDLVAQAGYQCTVLPVLVAVSWILATIEKFL 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 LTHPALTNANGVAAGRHTM-----NFGFEIAMIIGYQGVFPVLLAWENSIVEKQL 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 HKRLKGTADFLTPVLTLLLTGELFTFTAIIPANRWGVGVLAHGLQGLYDGPVGGLLFG 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 RRAIPDALDILTPFTVLIISGFTALLIIPAGRALGDGSGFVLSTLISHAGWLAGLLFG 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 120 LVYSPIVITGLHOSFPPIELELFNQ---GGSFFATASMANIAQGAACLAFFVLAKEKL 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 GLYSVIVITGIHSHFAVEAGLGNPSIGNVFLPLTIWANAANVAQGAACLAFFVKTDAKI 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 177 KGLAGASGVAVLGITPEAIFGVNLRWRPFFIGTAAIGGALIALFNKIKAVAGAAFG 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 KATLPSAFSAMLGITAAIFGLNLRVKEFFIAALIGGAAGGAWVSVHVYMTAVGLTAI 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 237 LGVVSIDAPQVMFLCAVVVTFIARCAAI 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 418 PGMAIVQASSLLNIIGMVIAGVAFVTSL 447

RESULT 11  
PTSB\_KLEPN STANDARD; PRT; 456 AA.

ID PTSB\_KLEPN STANDARD; PRT; 456 AA.  
AC F27219;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-permease IIBC component) (Phosphotransferase enzyme II, BC component) (EC 2.7.1.69) (EII-SCR).  
DE DE SCRA.  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]  
R1 SEQUENCE FROM N.A.  
R2 STRAIN=1033-SP14 / KAY2026;  
R3 MEDLINE=91312133; PubMed=1649946;  
RA Schmid K., Ebner R., Jahreis K., Lengeler J.W., Titgemeyer F.;  
RT "A sugar-specific porin, SCRy, is involved in sucrose uptake in enteric bacteria";  
RL Mol. Microbiol. 5:941-950(1991).

-I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A COVALENTLY BOUND EIIA DOMAIN. INSTEAD, EII-SCR-MEDIATED PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.

-I- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein histidine + sugar phosphate.

-I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.

-I- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

-I- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

-----  
EMBL: X57401; CAA040658.1; -.  
PIR: S15195; S15195.  
DR HSP; P05053; IIBA.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Transmembrane; Inner membrane; Phosphorylation.  
FT DOMAIN 1 ? EIIB DOMAIN.  
FT FT FT 456 EIIIC DOMAIN.  
FT FT FT 132 POTENTIAL.  
FT FT FT 144 POTENTIAL.  
FT FT FT 181 POTENTIAL.  
FT FT FT 209 POTENTIAL.  
FT FT FT 247 POTENTIAL.  
FT FT FT 267 POTENTIAL.  
FT FT FT 308 POTENTIAL.  
FT FT FT 330 POTENTIAL.  
FT FT FT 350 POTENTIAL.  
FT FT FT 360 POTENTIAL.  
FT FT FT 388 POTENTIAL.  
FT FT FT 428 POTENTIAL.  
FT MOD RES 26 226  
FT PHOSPHORYLATION (BY SIMILARITY)

FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 456 AA; 48022 MW; 4AFDF5405CAEFC66 CRC64;

Query Match 20.7%; Score 489; DB 1; Length 456;  
Best Local Similarity 38.59; Pred. No. 2.8e-25;  
Matches 104; Conservative 51; Mismatches 101; Indels 14; Gaps 3;  
QY 3 MVFPLSVNGYDVA---TMAAGEMPMSLGLDVAAGYQGTVLPLVSVWILATIERKFL 59  
DB 186 LTHPALTNAGVAAAGFHTM-----NFFGLEVAMICYQGTVPVLLAVFMFSMVERKL 237  
QY 60 HRLKGTADFLTPVLTLLGLTFIAIGPAMRWGVDLAHGLGGLYDFGPGVGLLFG 119  
DB 238 RRVIPDLDLITLPTLVIIISGFIALLLIGPAGRALGDGIFSLTSLHAGWLAGLLFG 297  
QY 120 LVYSPVITGLHOSPPPTLELFNQ---GGSFIFATASWANTAOCAAGLAVFFELAKSEKL 176  
DB 298 GLYSVIVITGHHSPHAEAGLLGNPSGVNFFLPIDWAMNVAQGGACFAVWEKTKDAKI 357  
QY 177 KGLAGAGSVAVLGITEPAIFGVNLRWPPFFIGIGTAAIGGALIALFNKAVKALGAAGF 236  
DB 358 KAITLPSAFSAMLGITEAIFGINLRVKPFTALVGVGAAGGAVVSMHVMTAVGLTAI 417  
QY 237 LGVNSIDAPDMVFLVCAVVFIFAGAAI 266  
DB 418 PGMATVQASLLNYITIGMAIAFAVAFALS 447

## RESULT 12

PTGA\_CORGL STANDARD; PRT; 674 AA.  
AC Q45298;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PTS system, glucose-specific IIBC component (EIIABC-GLC) (Glucose-  
DE permease IIBC component) (Phosphotransferase enzyme II, ABC  
DE component) (BC 2.7.1.69) (EII-GLC/EIIC-GLC).  
GN Corynebacterium glutamicum (Brevibacterium flavum).  
OG plasmid pBSB2.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13869;  
RA Yoon K.-H.;  
RT "Cloning and nucleotide sequence of enzyme II of Brevibacterium  
RT lactofermentum phosphotransferase system.";  
RL submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
CC histidine + sugar phosphate.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----  
CC EMBL; L18875; AAA22992.1; -.  
CC InterPro; IPR001127; PTS\_EIIA.  
CC InterPro; IPR001996; PTS\_EIIB.  
CC InterPro; IPR003352; PTS\_EIIC.  
CC Pfam; PF00358; PTS\_EIIA.1; 1.  
CC Pfam; PF00367; PTS\_EIIB; 1.  
CC Pfam; PF02378; PTS\_EIIC; 1.  
CC ProDom; PD001476; PTS\_EIIB; 1.  
CC ProDom; PD002243; PTS\_EIIA; 1.  
CC ProSITE; PS00371; PTS\_EIIA.1; 1.  
CC ProSITE; PS01035; PTS\_EIIB\_CYS; 1.  
CC KW phosphotransferase system; Sugar transport; Transferase;  
KW Phosphorylation; Transmembrane; plasmid.  
FT DOMAIN 1 43 EIIB DOMAIN.  
FT DOMAIN ? ? EIIC DOMAIN.  
FT DOMAIN 542 674 EIIA DOMAIN.  
FT TRANSMEM 126 146 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 193 213 POTENTIAL.  
FT TRANSMEM 225 245 POTENTIAL.  
FT TRANSMEM 260 280 POTENTIAL.  
FT TRANSMEM 303 323 POTENTIAL.  
FT TRANSMEM 344 364 POTENTIAL.  
FT TRANSMEM 377 397 POTENTIAL.  
FT TRANSMEM 409 429 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
FT TRANSMEM 493 513 POTENTIAL.  
FT MOD\_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 19.6%; Score 462.5; DB 1; Length 674;  
Best Local Similarity 30.0%; Pred. No. 2.4e-23;  
Matches 139; Conservative 79; Mismatches 204; Indels 41; Gaps 9;

QY 27 SLFGLDVAQAQYQGTVLPLVSVWILATIERKHLKRLGTADFLTPVLTLLTGFLTFI 86  
DB 218 TVFGLPMVINDYSGQVFPPLIAAGLVWVERALKRIPEAVQMVFPFSLIMIPATAF 277  
QY 87 AIGPAMRWGVDLAHGLGGLYDFGPGVGLLFGVSPVITGLHQSPPPTLELFNQ-G 145  
DB 278 LLGPFGLGVNGVGISSLEAVNFSFFILSVIPLPLVPLGLHWPPLNAINQNLTLG 337  
QY 146 GSFIPTASMANIAQGAACLAFFFLAKSEKLUKLAGAS--GVSAVL--GITEPAIFGVNL 201  
DB 338 YDFIQGPMGAMNFACFGLVTVGFVLIALKENRAMRQVSLGGLAGLLGGISEPSLYGVLL 397  
QY 202 RLRWPEFIGTAAIGGALIALFNKAVKALGAAGLGVVSDADPMVFLVCAVVFIFIA 261  
DB 398 RFKTYTFLPLGCVLGGVGMGIFDIKAYAF---VFTSLTLPAMDPMWLGTYTVGTAAPV 452  
QY 262 FGAAIAYGLYLVRNGSID-----PDATAAPV-----PAGTTKA 295  
DB 453 FTSMLLVLPDYRSDAERDEAKAQAARQNTPTPAAPVAPAAAGAAAGGAGAT-A 511  
QY 296 EAEAPAEFSDSTIIQAPLTGEATLSVSDAMFASGKLGSVAIVPTKGLGVSPVSGKI 355  
DB 512 VATKPLAAGOLVEITSPLEGHAVPLSEVPDPIFAAGKLGPGIAIEPTGTWVAPADATV 571  
QY 356 VVAPPSGHAFVTKAEDGSNVDTLMHIGFDTVNLNGTHFNPLKQGVKAGELLCEFD 415  
DB 572 ILVKQSGHVALRUE----SGVELLIHIGLDTVLQVGEGEVKRVHVKQVAGDPLITFD 627  
QY 416 IDAIIKAAGYVTTPIVSVNKKTKGTVNTYGLGTEAGANLNV 458  
DB 628 PEFRSRKNLPLITPVVSNANKFGEIVGIEAAQADATTVIKV 670

RESULT 13

SACX\_BACSU  
ID SACX\_BACSU STANDARD; PRT; 459 AA.  
AC P15400;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Negative regulatory protein of sacY.  
OS SACX OR SACS OR IPA-14R.  
GN Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90337338; PubMed=2116367;  
RA Zukowski M.M., Miller L., Cogswell P., Chen K., Aymerich S.,  
RA Steinmetz M.;  
RT "Nucleotide sequence of the sacS locus of Bacillus subtilis reveals  
the presence of two regulatory genes.";  
RL Gene 90:153-155(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95020537; PubMed=7934828;  
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
RA Rapoport G., Danchin A.;  
RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
kb region from 325 degrees to 333 degrees.";  
RL Mol. Microbiol. 10:371-384(1993).  
RN [3]  
RP SEQUENCE OF 348-459 FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=92216127; PubMed=1806041;  
RA Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A.,  
RA Dedonder R.;  
RT "A gene encoding a tyrosine tRNA synthetase is located near sacS in  
Bacillus subtilis.";  
RL DNA Seq. 1:251-261(1991).  
CC -!- FUNCTION: Negatively regulates sacY.  
CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M29333; AAA75335.1; -  
DR EMBL; X52480; CAA36719.1; -  
DR EMBL; X73124; CAA51570.1; -  
DR EMBL; J99123; CAB15867.1; -  
DR PIR; J00293; J00293.  
DR PIR; S16421; S16421.  
DR HSSP; P05053; IIBA.  
DR Subtilist; BG10560; sacY.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF03378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR ProSITE; PS01035; PTS\_EIIB\_CYS; 1.  
KW Transcription regulation; Transference; Phosphorylation;  
FT Complete proteome.  
FT DOMAIN 1 ? EIIB DOMAIN.  
FT DOMAIN ? 459 EIIC DOMAIN.  
FT MOD\_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 136 136 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 433 433 PHOSPHORYLATION (BY SIMILARITY).

SQ SEQUENCE 459 AA; 49024 MW; A5C4E996ECDA3D40 CRC64;  
Query Match 17.2%; Score 406; DB 1; Length 459;  
Best Local Similarity 37.9%; Pred. No. 8.3e-20;  
Matches 96; Conservative 47; Mismatches 88; Indels 22; Caps 6;  
QY 28 LFGLDVAQAQYQGTVPVLVSVWILATIEKFLHKLKGTADFLITPVLTLTLTGELTFIA 87  
Db 206 IWGLHIPMMGYQGMIPILLVSFVMSKIEKLLKSIYVPSKLDVVYIFITVWVVGCLALIV 265  
QY 88 IGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFGLVSPVITVGLHQSPFPIELELF---NQ 144  
Db 266 MNPAASTIGQIMTQISVIYDHAGIAAGALFGGIYSTIVLSGLHSHFYATEATLLANPHV 325  
QY 145 GGSFIFATASMANIAOAGACLAFFFLAKSEKLLAGASCVSALVITEPAIFGVNRLR 204  
Db 326 GVNFLVPIWMSMANVAGGAGLAVFLTKQSSLLKIALPASLTAFGLGIVEIVFGVNLKLI 385  
QY 205 WPEFEGIGTA---AIGGALIALFNKAVAGALGAGLGVVSDAP-----DMVMFLVC 253  
Db 386 RPF---IGAAIGAGIAGYVAVOVVANSYGLTG-IPMISIVLPFGAANFVHYMIGFLIA 441  
QY 254 AV---VTFEIAF 262  
Db 442 AVSAFIATLFLGF 454  
RESULT 14  
ID PTSB\_BACSU STANDARD; PRT; 460 AA.  
AC P05306;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-  
DE permease IIBC component) (Phosphotransferase enzyme II, BC component)  
DE (EC 2.7.1.69) (EIIC-SCR).  
GN SACP OR IPA-49D.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88097369; PubMed=3122206;  
RA Fouet A., Arnaud M., Klier A., Rapoport G.;  
RT "Bacillus subtilis sucrose-specific enzyme II of the  
RT phosphotransferase system: expression in Escherichia coli and  
RT homology to enzymes II from enteric bacteria.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95020537; PubMed=7934828;  
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
RA Rapoport G., Danchin A.;  
RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
RT kb region from 325 degrees to 333 degrees.";  
RL Mol. Microbiol. 10:371-384(1993).  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY  
CC SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: protein N-phosphohistidine + sugar = protein  
CC histidine + sugar phosphate.



FT	MOD_RES	29	29	PHOSPHORYLATION (BY SIMILARITY).
FT	TRANSMEM	111	131	POTENTIAL.
FT	TRANSMEM	159	179	POTENTIAL.
FT	TRANSMEM	189	209	POTENTIAL.
FT	TRANSMEM	226	246	POTENTIAL.
FT	TRANSMEM	259	279	POTENTIAL.
FT	TRANSMEM	301	321	POTENTIAL.
FT	TRANSMEM	341	361	POTENTIAL.
FT	TRANSMEM	371	391	POTENTIAL.
FT	TRANSMEM	399	423	POTENTIAL.
FT	TRANSMEM	441	461	POTENTIAL.
FT	CONFLICT	126	127	GL -> PF (IN REF. 1).
FT	CONFLICT	184	194	GCTPLIGVLG -> AORSLVSLA (IN REF. 2).
FT	CONFLICT	187	187	MISSING (IN REF. 1).
FT	CONFLICT	307	307	A -> Q (IN REF. 1).
FT	CONFLICT	429	473	FGILSIQPSYQWQFALAMATAIIIPVLTSTIYQRYKRLGT
FT				LDIV -> RNSLSTELLAGVACNGYRHHHPDCTHLVYLS
FT				AEIPPGHAGHCLIFFGAQLRSHQE (IN REF. 1).
SQ	SEQUENCE	473 AA;	51080 MW;	7437F8822B624944 CRC64;
Query Match 15.5%; Score 367; DB 1; Length 473;				
Best Local Similarity 30.1%; Pred. No. 3.le-17;				
Matches 80; Conservative 63; Mismatches 109; Indels 14; Gaps 5;				
QY	1	MAMVFP	SLVNGYDVAATMAAGEMP-MWSL	FGLDVAQAGYQGTVPVLPVSVWILATIEKFL 59
DB	195	VTLV	SQLMNAI-----LIGQQL	PEVDFGMFSIAKVGQAQVIFALLAGLALGVETRL 249
QY	60	HKRL	AGTADF	LTPVLTLLLTGFTFIATGPAMRWGDVLA----HGLQGLYDFGGPVGG 115
DB	250	KRIV	PDYLVVVV	VCSLILAVFLAHALIGPGRMIGDGVAFVRHMTGSP---APIGA 306
QY	116	LLFGL	VYSPVITGLHQSPPIEL	ELF-NQGSFTIFATASMANIAQGAACLAVFFLAKSE 174
DB	307	ALF	GFYAPLVITGVHTT	LAIIDLQIOSMGCTPVWPLIALSNIAQGSAVIGIITSSRKH 366
QY	175	KLK	GLAGAGSV	AVLGITEPAIFGVNLRLRWPFFIGTGTAAIGGALLALENIKAVALGAA 234
DB	367	NERE	ISVPAASAWL	GVTEPAMYGINLKYRFPMLCAMIGSLGLLCCGLNGVMANGIGVG 426
QY	235	GFL	GVVSIDAPDM	VMFLCAVVTFET 260
DB	427	GLP	GILSIQPSY	WQVQFALAMATAIIII 452

Search completed: September 25, 2002, 06:17:42  
Job time: 998 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:58:54 ; Search time 71.46 Seconds  
(without alignments)  
1132.964 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVPSLVNGYDVAAATMAA.....TEAGANLLNVAKEAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19.\*

- 1: sp.archaea.\*
- 2: sp.bacteria.\*
- 3: sp.fungi.\*
- 4: sp.human.\*
- 5: sp.invertebrate.\*
- 6: sp.mammal.\*
- 7: sp.mhc.\*
- 8: sp.organelle.\*
- 9: sp.phage.\*
- 10: sp.plant.\*
- 11: sp.protist.\*
- 12: sp.virus.\*
- 13: sp.vertebrate.\*
- 14: sp.unclassified.\*
- 15: sp.rviro.\*
- 16: sp.bacteriophage.\*
- 17: sp.archae.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	977	41.3	534	2	Q93ML1	Q93ml1 lactobacill
2	962	40.7	650	2	Q9S6S6	Q9s6s6 lactococcus
3	759.5	32.1	627	16	Q9L8G6	Q9l8g6 clostridium
4	695.5	29.4	630	16	Q9KG19	Q9kg19 bacillus ha
5	674	28.5	633	16	Q92FS7	Q92fs7 listeria in
6	673.5	28.5	636	16	Q9KF90	Q9kfi90 bacillus ha
7	673	28.5	655	16	Q97NW9	Q97nw9 streptococc
8	664	28.1	627	16	Q97PB8	Q97pb8 streptococc
9	659	27.9	479	16	Q9KLT8	Q9klt8 vibrio chol
10	659	27.9	480	16	Q99R00	Q99rq0 staphylococ
11	658.5	27.9	644	2	Q9KJ80	Q9k380 streptococc
12	647.5	27.4	674	16	Q99XQ4	Q99xq4 streptococc
13	632	26.7	620	16	Q99Y91	Q99y91 streptococc
14	612.5	25.9	621	2	Q48A08	Q48a08 klebsiella
15	612	25.9	628	16	Q97J79	Q97j79 clostridium
16	607	25.7	577	2	Q91461	Q91461 lactobacill

17	589.5	24.9	616	2	Q46129	Q46129 clostridium
18	588.5	24.9	636	16	Q9CFK9	Q9cfk9 lactococcus
19	570.5	24.1	612	16	Q97S38	Q97s38 streptococc
20	569.5	24.1	620	16	Q9KAS1	Q9kas1 bacillus ha
21	564.5	23.9	620	16	Q9A0X4	Q9a0x4 streptococc
22	530.5	22.5	618	2	Q9X565	Q9x565 enterococcu
23	517.5	21.9	480	2	Q9F8X3	Q9f8x3 pseudomonas
24	509	21.5	640	2	Q9RLJ2	Q9rlu2 listeria mo
25	507	21.5	661	2	Q68468	Q68468 corynebacte
26	499	21.1	457	2	Q9F499	Q9fi99 erwinia amy
27	487	20.6	475	16	Q99WC9	Q99wc9 staphylococ
28	487	20.6	458	16	Q9KBR9	Q9kbr9 bacillus ha
29	468.5	19.8	683	2	Q46072	Q46072 corynebacte
30	467	19.8	372	2	Q45622	Q45622 bacillus sp
31	462.5	19.6	681	2	Q93WZ4	Q93wz4 corynebacte
32	458.5	19.4	465	2	P94470	P94470 bacillus st
33	396.5	16.8	494	16	Q92CE9	Q92ce9 listeria in
34	388	16.4	451	2	Q9ZHU9	Q9zh9 clostridium
35	348.5	14.7	681	16	Q99X32	Q99x32 staphylococ
36	345	14.6	665	16	Q97LJ0	Q97lj0 clostridium
37	342.5	14.5	453	16	Q9K700	Q9k700 bacillus ha
38	341.5	14.5	692	2	Q53922	Q53922 staphylococ
39	339	14.3	478	16	Q9KTJ2	Q9ktj2 vibrio chol
40	323	13.7	484	16	Q99X29	Q99x29 staphylococ
41	314	13.3	195	16	Q9CGG4	Q9cgg4 lactococcus
42	307	13.0	652	2	Q938Q7	Q938q7 lactobacill
43	305.5	12.9	675	16	Q9KEK8	Q9kek8 bacillus ha
44	294	12.4	688	16	Q99R97	Q99r97 staphylococ
45	289.5	12.3	675	2	Q57071	Q57071 staphylococ

ALIGNMENTS

RESULT 1

Q93ML1	PRELIMINARY;	PRT;	534 AA.
ID	Q93ML1		
AC	Q93ML1;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	SUCROSE-SPECIFIC ENZYME II OF THE PTS (FRAGMENT).		
GN	SCRA.		
OS	Lactobacillus sakei.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;		
OC	Lactobacillus.		
OX	NCBI_TaxID=1599;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Dudez A.-M., Chailou S., Hissler L., Stentz R., Champomier-Verges M.,		
RA	Alpert C.-A., Zagorec M.;		
RT	*Physical and genetic map of the Lactobacillus sakei 23K chromosome.*;		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF401046; AAK92528.1; "		
FT	NON_TER 1		
SQ	SEQUENCE 534 AA; 56230 MW; 01804F9DE70C0089 CRC64;		

Query Match	41.3%;	Score 977;	DB 2;	Length 534;
Best Local Similarity	44.2%;	Pred. No. 3.5e-53;		
Matches 211;	Conservative 90;	Mismatches 138;	Indels 38;	Gaps 10;
QY	1	MAMVPSLVNGYDVAAATMAAGEMPMSLFG	LDDAQAQYQGTVP	LPVLVYVSWILATIERFLH 60
Db	81	MMLVMPNLVNGYGVASIGTGHMTYHVFGL	NIAQAQYQGTVP	IVIGVAFILANLEKFFH 140
QY	61	KRLKGTADFLITPVLLTLGLTFTTAIG	PAMRWGVDVLAHGLQGLYDFGPGVGLLFL 120	
Db	141	KHLNDADVDTFTPLMSIIITGFTLT	VLGPAIRVNSGVTDSLWYAYQTILGAVGMIFGL 200	
QY	121	VYSPIVTGLHQSPPTLELFL----	NOGGSFIFATASMANIAOGAACLAVFFLAKSEKL 176	
Db	201	GYSAILVTGLHQSPFAETETLLADI	AKTGTSFIPVAAANIAOGAATFAVFFVTNNKQ 260	

QY 177 KGLAGASGVAVLGTTEPAIFGVNLRWPPFFIGTAAIGGALIALFNKAVAGRAF 236  
 Db 261 KSLTTSAGISAMLGITTEPALFGVNLKLFPPFIIGLHVLVLSVSGMPAGI 320  
 QY 237 LGVVSIDAPDMVFLVCAVVTFFIAFGAAIAJGLVLRNGSIDPDATAAPVPAGTTKAE 296  
 Db 321 IGFTAIAPKSPSPWKGAIISFVAFVGTLYLKGKAMK-----TTEEE 363  
 QY 297 A--EAPA-----EFSNDSTIIQ-----APLTGBAIALSSVSDFAMFASGKLGSGVAIVPTKGLVS 349  
 Db 364 IINEAPATPEVVERLQDEK-ISPATGRIVDLASVPDPVFSEAKMGKGIATMPTSQDVL 422  
 QY 350 PVSQKIVVAFPSGHAFVTRKAEDGSNDVILMHIGFDVNLNGHTFNPPLKQGDDEVK 409  
 Db 423 PVTGVITTAANTGHAYGT--KSDGGA--EVLHIGLDFVNLNGIGTFEKIVQOQGVHVS 478  
 QY 410 LCLCFEFDIDAIIKAAGVEVTTPIVWSN---YKKTGPVNTYGLGEIEAGANLLNV-AKKE 462  
 Db 479 LLGHFEDIDIKIQAGLPTLTWIVNTAGYAQVDFLLTVDKAAMQ-GEETIQLHAKKD 534

## RESULT 2

Q9S6S6 PRELIMINARY; PRT; 650 AA.  
 AC Q9S6S6  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-JUN-2001 (TREMBlrel. 13, Last sequence update)  
 DE ENZYME II SUCROSE PROTEIN (EC 2.7.1.69).  
 GN SACH.  
 OS Lactococcus lactis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99173919; PubMed=10074089;  
 RA Luesink E.J., Maruy J.D., Kuipers O.P., De Vos W.M.;  
 RT "Characterization of the divergent sacB and sacAR operons, involved  
 in sucrose utilization by Lactococcus lactis.";  
 RL J. Bacteriol. 181:1924-1926(1999).  
 DR EMBL; Z97015; CAB09690.1; -  
 DR HSP; P45618; ZGPR.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Transferase.  
 SQ SEQUENCE 650 AA; 69636 MW; 250E26F7664D4204 CRC64;

Query Match 40.7%; Score 962; DB 2; Length 650;  
 Best Local Similarity 42.9%; Pred. No. 3.8e-52;  
 Matches 203; Conservative 89; Mismatches 145; Indels 36; Gaps 7;

QY 1 MAMVFPSPVNGYDVAAATWAGEMPWSLFLGLDVAQAQYQGTVLVPLVVSWTIAIEKFLH 60  
 Db 194 MIMVMPGLINGYNVAEASNNHTMTYDFIGFKVAQAQYQGVLPVIGVAFIAKLERFFH 253  
 QY 61 KRLGTADFLLTPTLLTGLTFLFIAIGPAMRWGVDLAHGLQGLYDFGSPVGLLFLGL 120  
 Db 254 KYLNDADIFTFTPLLSVITGTLFTTIVGPAIRFVSNGLDGLVGLNLTGALGNLVEGG 313  
 QY 121 VYSPVITGLHQSPFPPIELELFNQ-----GGSPFIFATASMANIAQAACLAFFLAKSE 174  
 Db 314 FYSAIVVTGLHQSPFAIETMLITNYQHSIGGIDGFFPVAACANNAQAGATFAILFTVTKNI 373

QY 175 KKLGLAGASGVSNVLGTTEPAIFGVNLRWPPFFIGTAAIGGALIALFNKAVAGRAF 234  
 Db 374 KTKALAAPAGVSAIILGITTEPALFGVNLKLFPPFIIGLHVLVLSVSGMPAGI 433  
 QY 235 GFLGVVSIDAPDMVFLVCAVVTFFIAFGAAIAJGLVLRNGSIDPDATAAPVPAGTTK 294  
 Db 434 GLIFISIKAGYNLQFMISIFISFLIAFVTVSYGRRMEAK--SITKE-----K 480  
 QY 295 AEAPAPAEFSNDSTIIQ-----APLTGBAIALSSVSDFAMFASGKLGSGVAIVPTKGG 346  
 Db 481 NKQNTATQYQPEKVLIDPVKSGELLAPINGFVPLSDVSDPVFSKEIMGKGIATPKSGE 540  
 QY 347 LVSPVSGKIVVAFPSGHAFVTRKAEDGSNDVILMHIGFDVNLNGHTFNPPLKQGDDEVK 406  
 Db 541 LFPSPADGEIITAYTGHAYGKTK----NGGEVLHIGIDTVSMNGNGFIQNVKVGQKVK 596  
 QY 407 AGEILCFEFDIDAIIKAAGVEVTTPIVWSN---YKKTGPVNTYGLGEIEAGANLL 456  
 Db 597 AGDLLSGFDEKKEIKSGLDVTVIIVTNSASYNELPLSEN--VDIKVGEKIL 647

## RESULT 3

Q9L8G6 PRELIMINARY; PRT; 627 AA.  
 AC Q9L8G6  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SCRA (FUSION: PTS SYSTEM, BETA-GLUCOSIDES SPECIFIC IIABC COMPONENT).  
 GN SCRA OR CAC0423.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE-20391269; PubMed=10937490;  
 RA Tangney M., Mitchell W.J.;  
 RT "Analysis of a catabolic operon for sucrose transport and metabolism  
 in Clostridium acetobutylicum ATCC 824.";  
 RL J. Mol. Microbiol. Biotechnol. 2:71-80(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE-21359325; PubMed=11456286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Soucaille P., Daly M.J.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Smith D.R.;  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AF205034; AAF35839.1; -  
 DR EMBL; AE007557; AAK78403.1; -  
 DR HSP; P20166; IGPR.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR PROSITE; PS00371; PTS\_EIIA\_1; UNKNOWN\_1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 627 AA; 67596 MW; B835AB0238FCA436 CRC64;

Query Match 32.1%; Score 759.5; DB 16; Length 627;  
 Best Local Similarity 36.7%; Pred. No. 1.5e-39;

		Matches 163; Conservative 80; Mismatches 176; Indels 25; Gaps 5;
QY	3	MYEPLVNGYDYDAATMAAGEMPWSLFGLDVAAGYQGTVPVLVSVWILATIEKFLHR 62
Db	186	MTHPDLQNAW---TLGEGIKHTITFGLNIGMVGYGQGTVPILISVWMSYIEKGLRK 241
QY	63	LKGTADFLTPVLTLLTGLTFIAIGPAMRWGDVLAHGLGLOGLYDFGPGVGLFLGLV 122
Db	242	VPEALDILLTPFLTMITGFFAMVIGPGRFVGDEISLGLOTLYNTGFFSGVLEGGY 301
QY	123	SPVITGLHQSPPPIELEFNQGG---SFIFATASMANIAQAACIATVFFLAKSEKLG 179
Db	302	SLIVITGIHSEFAIEAGLANPAIHKFLLPWSNANVAAGGAALAVYFKTRDKKMSI 361
QY	180	AGASGSAVLGITEPAIFGVNLRWPFIFIGTAAIGALIALFNKAVAGLAFV 239
Db	362	APASESCLGLITEPAIFGVNLRYPKFTAGALGGAGGYYVFTKAVMTAVGVTGIPGI 421
QY	240	VSIDAPDMVFLVCVVTFEIAFGAAIAYGLVLRNRSIDPDATAAPVPAGTTKAEAA 299
Db	422	AIVKQGSFLNYIAMI---LAFGGFIANVLGIKEEITEEDLNKETVNDKIKVEVES 477
QY	300	PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSAIVPTKGQLVSPVSGKIVVAF 359
Db	478	-----VVSPVNGKVLLKNVPDKTFAEGLIGDGVDPEDGCVVSPIDGTVVHVF 527
QY	360	PSGHAFVATKAEKDSNVDIILMHIGFDTVNLNGTHENPLKKGDEVKAGELICEFDIDAI 419
Db	528	ETKHAJAMSKS---NGVEMLIHIGDITVMEGNGFKSFINDGEEYKKGDKLIQFDLDLV 583
QY	420	KAAGVEVTPPIVSVNYKTKGPVNT 443
Db	584	KEKAVSPILVTIVNHEDMGFVNS 607
RESULT	4	
Q9KG19		
ID	Q9KG19	PRELIMINARY; PRT; 630 AA.
AC	Q9KG19	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.	
GN	BH0296	
OS	Bacillus halodurans.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus.	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	*Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis.;	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
DR	EMBL; AP001508; BAB04015.1; -.	
DR	HSSP; P20166; 1AX3.	
DR	InterPro; IPR001127; PTS_EIHA.	
DR	InterPro; IPR001996; PTS_EIIB.	
DR	InterPro; IPR003352; PTS_EIIC.	
DR	Pfam; PF00358; PTS_EIHA_1; 1.	
DR	Pfam; PF00367; PTS_EIIB_1; 1.	
DR	Pfam; PF02378; PTS_EIIC; 1.	
DR	ProDom; PD002243; PTS_EIHA; 1.	
DR	PROSITE; PS00371; PTS_EIHA_1; 1.	
DR	PROSITE; PS01035; PTS_EIIB_Cis; 1.	
KW	Complete proteome.	
SEQUENCE	630 AA; 66978 MW; 6F0218011686ADD5 CRC64;	

		Query Match 29.4%; Score 695.5; DB 16; Length 630;
		Best Local Similarity 32.8%; Pred. No. 1.5e-35;
		Matches 152; Conservative 108; Mismatches 182; Indels 21; Gaps 5;
QY	2	AMVFPSLVNGYDYDAATMAAGEMPWSLFGLDVAAGYQGTVPVLVSVWILATIEKFLHK 61
Db	185	ALIYPSIVELHDSAIDV-----TFFGIPVVMYNTSVFFILLAVFAMSIVEKCNK 236
QY	62	RKGTADFLTPVLTLLTGLTFIAIGPAMRWGDVLAHGLGLOGLYDFGPGVGLFLGLV 121
Db	237	KIHEAVKNFVTPILLVWVVPVTLIILGPIGVYLGNGIASVQIEFTFSPVLGAIVAGI 296
QY	122	YSPVITGLHQSPPPIELEFNQGGSFIFATASMANIAQAACIATVFFLAKSEKLG 180
Db	297	MQVLVIFGIHWGIIPIILNLSVRGEDVIKAAAPAVFSQAGALGVMLRTRKKLKALA 356
QY	181	GASGSAVLGITEPAIFGVNLRWPFIFIGTAAIGALIALFNKAVAGLAFV 240
Db	357	GSTSITALGITEPAVVGVTPLPKKPFIMAVISAAGVAIGHVGSVAVAPGAPLLTIP 416
QY	241	SI---DAPDMVFLVCVVTFEIAFGAAIAYGLVLRNRSIDPDATAAPVPAGTTKAE 297
Db	417	IFYPEDGRGFVAFVIAIISFVLA----AVLTYYIVGFKDPVDEDTLSNESGSENEVKR 471
QY	298	EAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSAIVPTKGQLVSPVSGKIVV 357
Db	472	EDDKKPSSEEEKSLKGEVPLTEVQDHVFSGGAMGKGVAVRPREGRVUAVPINTVTS 531
QY	358	AFPSGHAFVATKAEKDSNVDIILMHIGFDTVNLNGTHENPLKKGDEVKAGELICEFDID 417
Db	532	LPETKHAIGITS---DNGTEIFIHVIGDITVQLKGEHFTSFIEQGSDEVAAGDVLLEFDE 587
QY	418	ATKAAGYEVTPPIVSVNYKTKGPVNTYGLGEIEAGANLLNVAK 460
Db	588	RITAAGYDVTPIVLTNAKQFSNVQTTDKREVTSEDLIIHVIK 630
RESULT	5	
Q9ZFS7		
ID	Q9ZFS7	PRELIMINARY; PRT; 633 AA.
AC	Q9ZFS7	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	LIN0026 PROTEIN.	
GN	LIN0026	
OS	Listeria innocua.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Listeria.	
OX	NCBI_TaxID=1642;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CLIP 11262 / SEROVAR 6A;	
RX	PubMed=11679669;	
RA	Glaser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,	
RA	Baquerio F., Berche P., Bloeker H., Brandt P., Chakraborty T.,	
RA	Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,	
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,	
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,	
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,	
RA	Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,	
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,	
RA	Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,	
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,	
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;	
RT	*Comparative genomics of Listeria species.;	
RL	Science 294:849-852(2001).	
DR	EMBL; AL596163; CAC95259.1; -.	
DR	ListiList; LIN00026; -.	
KW	Complete proteome.	
SEQUENCE	633 AA; 67277 MW; 8AD80445D96C35A0 CRC64;	

Query Match 28.5%; Score 674; DB 16; Length 633;  
Best Local Similarity 36.2%; Pred. No. 3.3e-34;  
Matches 161; Conservative 75; Mismatches 179; Indels 30; Gaps 10;

QY 2 AMWPSSL-VNGYDVAATMAAG---EMPMSLF-GLDVAQAQYOGTVLPVLVSWILATI 55  
DB 181 ALVPTLAGISAGDPIYTLFAGTIFESPIHTFGLIPVILMSYASSVPIITLTYFGSKV 240

QY 56 EKFLHKLKRGADFLITPVLLTLLTGFATIAIGPAMRWGVDVLAHGLQGLYDFGPGVG 115  
DB 241 EKGFKKIIPDVKTFFVPECTLLVVVPITFLVIGPIATWAGOLLGAGTIWYNLSPPVAG 300

QY 116 LLFGLVSPVITGLHQSPPIEL-ELENQGSFIFATASMANIQAACLAFLAKSE 174  
DB 301 LILGFGWVVFIFGLHWGLIPIVAINNLTGLSDPVLAMFAGSQAQIGAVLAVFEKTRNK 360

QY 175 KLKLAGASGVSALVIGITEPAIFGNLRWRPFIFIGTAAIGGALIALENKAVAGAA 234  
DB 361 KIKLSIPAFISGIFGVTEPAIYGVTLPLKPPFIMSCITAGAVGGIIGFVSAKYIMGGL 420

QY 235 GFLGVVSDAP-----DMVMELVCAVTVFFIAFGAAIAYGLYLVRNRGSDIDATAAPV 289  
DB 421 GIFGLPNFPQAGITSFAFWVIAIVISFI-----LGFILTVVAGFKDP---AEAVV 470

QY 290 AGTTKAEAEAPAEBSNDSTIIQAPLTGEAIALSSVSDAMFASGLGSGVAIVPTKGQLVS 349  
DB 471 EETNVTEGETLIE----RETIPAPVWGIVTLADVKDEAFSSGALGKGAIVPTVGRVVA 526

QY 350 PVSQKIVVAPPSGHAFVTRKAEDGSNVDILMHIGEDFVNLTGTHFNPLKQGEVAKAGE 409  
DB 527 PAAGTVTIFTHAIGITT--NDGA--EVLHIGMDIVQLEGFTTAHVKGQDVIKQ 582

QY 410 LLCEFDIDAIAKAAGYEVTPVIVSN 434  
DB 583 LITEFDIEGIRAAGYVTPVIVSN 607

RESULT 6  
Q9KF90 ID Q9KF90 PRELIMINARY; PRT; 636 AA.  
AC Q9KF90;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.  
GN BGLP OR BH0595.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001509; BAB04314.1; -;  
DR HSSP; P45618; 2GPR.  
DR InterPro; IPR001127; PTS\_EIIA.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR ProDom; PD002243; PTS\_EIIA; 1.  
DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
KW Complete proteome.

QY 2 AMWPSSL-VNGYDVAATMAAGEMPMSLF-----GLDVAQAQYOGTVLPVLV 47  
DB 183 SLVPTLV-----VLTEGE-PLYTLTGTFESPVHTFGLIPVILMSYATSVIPIIL 234

QY 48 VSWILATIEKELHKLKGTADFLITPVLLTLLTGFATIAIGPAMRWGVDVLAHGLQGLY 107  
DB 235 AAYFASKVEARLRIIPDVVKTFVLPVFFTLVPLTIFVIGPIATWAGOLLGQFTLVVY 294

QY 108 DFGGPGVGLLGLVSPVITGLHQSPPIEL-ELENQGSFIFATASMANIQAACLA 166  
-DB 295 NLSPIIAGAFGGFQWQVFIQGLHWGLIPIAINNVUGSDPVLAMVFAAFAQIGAVAA 354

QY 167 VFFLAKSEKLKLAGASGVSALVIGITEPAIFGNLRWRPFIFIGTAAIGGALIALENI 226  
DB 355 VWLKIKQKQVKTLSVPAPFISGIFGVTEPAIYGVTLPLKPPFIISCIAAAGVGAIIGLPRS 414

QY 227 KAVALGAAGFLGVYSI-----DAPDMVMF-LYCAVTVFFIAFGAAIAYGLYLVRNRGSD 281  
DB 415 QGYIIGGLIGIFGIFSLHPADMDAGFWGIYAVVAVFV-LGFILTYLFLGLKSGNASDEQ 473

QY 282 DATAAPVAGTTPKAEAEAPAEBSNDSTIIQAPLTGEAIALSSVSDAMFASGLGSGVAIV 341  
DB 474 TETKAHSTGTGEKEE-----ISSPFNGSVITLSEIKDEAFSSGALGEGIAIE 521

QY 342 PTKQLVSPVSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGEDFVNLTGTHFNPLKQ 401  
DB 522 PSEGLFSPVSGMVTALYPTTHALGIT---DRGAELLIHGLDVTQLDCKFFTAHTIQ 577

QY 402 GDEVKAGELLCEFDIDAIAKAAGYEVTPVIVSNKTKGTPVNTYIGETEGAGANLVNAK 460  
DB 578 GAQVEKGDLLIEFDIKEIKAAGYAVTTPVIVTNHKOYQGLFDTDKQVNAQGRDLIELTR 636

RESULT 7  
Q97NW9 ID Q97NW9 PRELIMINARY; PRT; 655 AA.  
AC Q97NW9;  
DT 01-OCT-2001 (TremBLrel. 18, Created)  
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE TREHALOSE PTS SYSTEM, IIABC COMPONENTS.  
GN SPI884.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapfle E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL; AE007479; AAK75956.1; -;  
DR TIGR; SPI884; -;  
DR InterPro; IPR001127; PTS\_EIIA.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.

DR InterPro: IPR001950; SUI1.  
DR Pfam: PF00358; PTS\_EIIA.1; 1.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD002243; PTS\_EIIA; 1.  
DR PROSITE: PS00371; PTS\_EIIA.1; UNKNOWN\_1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; UNKNOWN\_1.  
DR PROSITE: PS01118; SUI1\_1; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 655 AA; 69785 MW; BD8C8C55C9DE0953 CRC64;

Query Match 28.5%; Score 673; DB 16; Length 655;  
Best Local Similarity 35.3%; Pred. No. 4e-34;  
Matches 170; Conservative 87; Mismatches 173; Indels 52; Gaps 16;

QY 1 MAMVFFSLVNGYDVATMAAGEMP--MWSLFLGDVAQAQYQGVTVLVLVSVWILATIEKF 58  
DB 195 ICLVSPQLLNAYAVASTPAADIAANVWVNFYFTVNRIGYQAQVIPALLAGLSLYLEIF 254

QY 59 LHKRLKGTADFLTPVLTLLTGLTFIAIGPAMRWVGDVLAHGLQGLYDFG--GPVG-- 114  
DB 255 WKHHPFVSMIFVPELSLIPALILAHVTLGP-----IGWTIGQLSSVVLGLTGPVKWL 310

QY 115 -GLLFLGLVSPVITVIGLHQSFPPIELELF--NOGGSFIFATASMANIAOAGACLAVFFLAK 172  
DB 311 FGAIFGALYAPVITGLHHTNAIDTQLIADAGGTALWPMIALSNIAQGSAYFAYFMHR 370

QY 173 -SEKLGLAGAGSVAVLGTETPAIFGVNLRWMPFFIGTAAIGGALIALFNKAVAL 231  
DB 371 HDREAREQSLPATISAYLGVTGFPALPGVNVKVIYYPVAGTGSALAGMLSVTENVTAASI 430

QY 232 GAAGFLGVVSDAPDMVF-----LVCNVTFETFAAGATAYGLLYVRNGSIDPDATAAP 287  
DB 431 GIGLPGIISIQPYMLPAGTMLVAIVVPMULTF-----FFRKAGLFTK----- 475

QY 288 VPAGTTKAAE-----APAEFSNDSTI-----IQAPLTGEAIALSSVSDAMFASGKLGS 337  
DB 476 -TEGDTNLQAEFAVQAEEAEFVNHEPVELTSVEIISPLTGQVKLSQATDPIFASGVYQG 534

QY 338 VAIVPTKQLVSPVSKIVVAPSGHAFVTRKAEDGSNDVILMHIGFTVNLNGTHFNP 337  
DB 535 LVIEPSQGLTSPVNGTVTLVFTPKHAI--VSDEG--VELLIHIGMDTVGLDGKGFES 590

QY 398 LKKQGDVRAKAGLCEFDIDATKAAGYEVTPIVSN---YKKTGPVNTYGLGEIAGAN 454  
DB 591 LVVQGDHVTGQQLIRFDMVIRKAGLVETPVIITNQDAYTATIP--GTYP--TIQAGAS 648

QY 455 LL 456  
DB 649 LM 650

RESULT 8  
Q97PB8 PRELIMINARY; PRT; 627 AA.  
AC Q97PB8  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DE PTS SYSTEM IABC COMPONENTS.  
GN SPI722.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,

RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., C.L.,  
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen E.K.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL: AE007465; AAK75799.1; -.  
DR TIGR: SPI722; -.  
DR InterPro: IPR001127; PTS\_EIIA.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR Pfam: PF00358; PTS\_EIIA.1; 1.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD002243; PTS\_EIIA; 1.  
DR PROSITE: PS00371; PTS\_EIIA.1; 1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.  
KW Complete proteome.  
SQ SEQUENCE 627 AA; 65683 MW; 9298CDD7B2E6B85F CRC64;

Query Match 28.1%; Score 664; DB 16; Length 627;  
Best Local Similarity 34.3%; Pred. No. 1.4e-33;  
Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;

QY 1 MAMVFFSLVNGYDVATMAAGEMPMSLFLGDVAQAQYQGVTVLVLVSVWILATIEKFL 60  
DB 182 MMLVSGSLPNAMAVA---QGGEVTAMFFGF-IPVVGLOGSVLPAPFIIVGAKFEKAVR 237

QY 61 KRKCTADFLITPVLTLLTGLTFTTAIGPAMRWVGDVLAHGLQGLYDFGPGVGLLFL 120  
DB 238 KVPDVIDLVTTFVTLVNSILGLFVIGPVFHVNYIILATKAILSMFPGLGFLIG 297

QY 121 VYSPVITVIGLHQSFPPIELELFNQGSFIF-ATASMANIAOAGACLAVFFLAKSEKLG 179  
DB 298 VHQLIIVSVGVHIFNLLEVLQLLAADHANPNAIITAMTAAGATVAVGVTKNPKLKL 357

QY 180 AGASGVAVLGTETPAIFGVNLRWMPFFIGTAAIGGALIALFNKAVALGAAGFLG- 238  
DB 358 APPAALSAGLGTETPAIFGVNLRWMPFFLFIAGAGGGLASTILAGTNGTITIPGT 417

QY 239 VVSDAPDMVFLVCAVTVFFIAFGAIAIAGLYLVRNGSIDPDATAAPVATTKAAE 298  
DB 418 MLTVGVNGQLPQYLLMVAVSFALGFALTYMFGY-----EDEYDATAAKRAEAEKEE 470

QY 299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSVAIVPTKQLVSPVSGKIV 357  
DB 471 VAPAAQLQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIIVKPSQGVVYAPADAESVI 528

QY 358 AFPSGHAFVTRKAEDGSNDVILMHIGFTVNLNGTHFNP LKQGDVRAKAGLCEFDID 417  
DB 529 APPTGHAFGLKTR---NGAEVLHIVGIDTVSMNGDGFETKVAQGNKVKAGDVLGTGDSN 584

QY 418 AIKAAGYEVTPIVSN---YKKTGPVNTYGLGEIAGANLNV 458  
DB 585 KIAAAGLDDTTMVIVTNGTDYASVAPVAT---GSVAKGDVAVIEV 625

RESULT 9  
Q9KLT8 PRELIMINARY; PRT; 479 AA.  
AC Q9KLT8  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.  
GN VCA0653.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]

```
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004395; AAF96554.1; -
DR TIGR; VCA0653; -
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 479 AA; 50292 MW; DC0E129229957E3C CRC64;

Query Match 27.9%; Score 659; DB 16; Length 479;
Best Local Similarity 49.3%; Pred. No. 2e-33;
Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

QY 1 MAMVPPSLVNGYDVAAATMAAGE-MPMWSLFGLDVAQAGYQGVTLPLVLSVWILATIEKFLH 60
Db 194 MLWHPDNLNGWGRGASVSGVTPTWNLGFEIKVGGVSLPLVLSAYILAKIENGLR 253

QY 61 KRLKGTADFLITPVLTLLTGLTFIAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFL 120
Db 254 KIVPSVDNLLTPLMAIFITGLFTVVGPLTRDVGFMGLDALNWLVDAGFVGGALEGF 313

QY 121 VYSPVITGLHQSPPIELF-----NOGGSFIFATASMANIAGAACLAVFFLA 176
Db 314 IYAPVITGMHHSFIAITQTLADIVITGGTFIPPIAAMNSIAGAAALAVGVMTKETL 373

QY 177 KGLACAGSVAVLGTETPAIFGVNLRWPFIFIGTAAIGGALIALFNKAVAGAGF 236
Db 374 KGVAIPSGVITALLGTEPAFMFGVNLKRLPFIAICGAAALASAFITLNVKAQALGA 433

QY 237 LGVVSIDAPDMVMFLVCAVVTFFAAGAAI 266
Db 434 PGIISNPQIGYINGMWAISFVAALTV 463

RESULT 10
Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIIC COMPONENT.
GN SCRA OR SA2167 OR SAV2377.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879; 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kishida S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
```

```
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43489.1; -
DR EMBL; AP003365; BAB58539.1; -
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR PRODOM; PD001476; PTS_EIIB; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;

Query Match 27.9%; Score 659; DB 16; Length 480;
Best Local Similarity 48.6%; Pred. No. 2e-33;
Matches 139; Conservative 53; Mismatches 82; Indels 12; Gaps 5;

QY 1 MAMVPPSLVNGYDVAAATMAAGE-MPMWSLFGLDVAQAGYQGVTLPLVLSVWILATIEKFL 59
Db 194 MLVHPSLMSAYDFPKAVEAGRAIPYDVFGHLHNGVGTQGVLPMLVAAVILASIEKGL 253

QY 60 HKRLKGTADFLITPVLTLLTGLTFIAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFG 119
Db 254 RKVPTVLDNLLTLLPSLIFITAFITFSFVGPITRQLGWLSGLTWLYEFGAIGGLIFG 313

QY 120 LVYSPVITGLHQSPPIELF-----NOGGSFIFATASMANIAGAACLAVFFLA 174
Db 314 LLYAPVITGMHHSFIAVETTLIADATKTGSGFIFPIATMSNVAQGGAAIAAFFIKQNK 373

QY 175 KKLGLACAGSVAVLGTETPAIFGVNLRWPFIFIGTAAIGGALIALFNKAVAGAA 234
Db 374 KLGVASAGSISALLGTEPAFMFGVNLKRLPFIAICGAAALASAFITLNVKAALUGTA 433

QY 235 GFLGVVSIDA--PDMVFLVCAVVTFFAAGAAIAYGLYLVRNGS 278
Db 434 GLPGFISINPVHAGLHVFGVGTISFII---AIVTVLILSKRAN 475

RESULT 11
Q9KJ80 PRELIMINARY; PRT; 644 AA.
AC Q9KJ80;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GLUCOSIDE-SPECIFIC EII PERMEASE.
GN BGLP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG8;
RX MEDLINE=20340959; PubMed=10878120;
RA Cote C.K., Cvitkovitch D., Bielweis A.S., Honeyman A.L.;
RT "A novel beta-glucoside-specific pts locus from Streptococcus mutans
RT that is not inhibited by glucose.";
RL Microbiology 146:1555-1563(2000).
DR EMBL; AF206272; AAF89975.1; -
DR HSP; P20166; IGPR
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIB; 1.
DR PRODOM; PD002243; PTS_EIIB; 1.
DR PROSITE; PS00371; PTS_EIIB_1; UNKNOWN_1.
SQ SEQUENCE 644 AA; 69282 MW; 723B7FBBDD2794EB CRC64;
```

	Query Match	27.9%	Score 658.5;	DB 2;	Length 644;
	Best Local Similarity	33.8%;	Pred. No. 3.le-33;		
	Matches 160;	Conservative	89;	Mismatches 196;	Indels 29; Gaps 13;
QY	2	AWPSPSLVNGYDV---	AATMAAGEM--PMWS-LFGLDV--AQAGYQGTVLPVLVVSHIL 52		
Db	185	ALVPNTAASTVKRHPLYTLFGTLLIVSPIYSTFTGPIIFIPASSYLQTLPVIVAIWAG 244			
QY	53	ATIEKFLHRLKGTADELITPVLTLLLTGTFTTAIGPAMRWGDVLAHLGGQLDFGCP 112			
Db	245	SKIEFFKKIIPDVVKVFVPPFFILLITVPLSVLIGVMSWASDLVGAIFTGIYGPNV 304			
QY	113	VGLLFLGVISPIVITGLHQSFPTIE-LELFNQQGSFIATASMANIAOQAACLAUVFLA 171			
Db	305	IYGVLGAMQVLMVMFLGHMLGPLLAILELKQPGCVILVATIAIC-FAQAGSLLNMIMRT 363			
QY	172	KSEKLKLAGASGVSNVLGITETPAIFGVNLRRLRPWPFFIGITGAIGGALLIALFNKAVAL 231			
Db	364	NENKVQLSIAPAFISAULFGVTPEAYIGITLPMRVPFIMTCVSGAISGAYLALFNVKMQVM 423			
QY	232	CAAGFLGVVSIDAPD----MYMFLVCVVTFIAFGAAIAYGL-YLVRRNSIDPDATAA 286			
Db	424	GGMGLEAIPSIDPKNSMILHFILAIAMNFVLGFVLTQFKIPPLYCEPTSDSDVDK 483			
QY	287	PVPAGTTKAEAEAPAEFSNDSTIIQAPLUTGBAIALSSVSDAMFASGKLGSGVAIVPTKGO 346			
Db	484	EPPVKELK-----EIKQE--LISSPLIGVKLENVPDEVFASGAMKGIAIDPDGDI 534			
QY	347	LVSYPGSKIIVAFPSGHAFVRTKAEDGSNVDDILMHIGFDVNLTNGHTUENPLKQGDVEK 406			
Db	535	WVAPTKEGTVLVFTTKAVGLURT--ENGA--EILIHGMDTVSLAGKGFKSFEVGEHDVE 590			
QY	407	AGELLCFDIDAIRKAAGVEVTPPIVWSNYKKTGPVNTYGLCEIEAGANLLNAVAK 460			
Db	591	AGQTLLEDVWNIAKAGLPVITPVIVINSQDFEDVLTQERTVEAGNVLITAVK 644			
RESULT	12				
Q99XQ4	ID	Q99XQ4	PRELIMINARY;	PRT;	674 AA.
AC	DC	Q99XQ4;	2001 (TEMBLrel. 17, Created)		
DT	DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)			
DT	DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)			
GN	DE	PUTATIVE PTS SYSTEM ENZYME II.			
OS	GN	SPY2097.			
OC	OS	Streptococcus pyogenes.			
OC	OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
RN	NCBI_Taxid=1314;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;				
RX	MEDLINE-21192684; PubMed-11296296;				
RA	Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,				
RA	Primeaux C., Sezate S., Suvoarov A.N., Kenton S., Lai H.S., Lin S.P.,				
RA	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,				
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;				
RT	*Complete genome sequence of an M1 strain of Streptococcus pyogenes.*;				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).				
DR	EMBL; AE006630; AAK34748.1; .				
DR	HSSP; P20166; IGPR.				
DR	InterPro; IPR001127; PTS_EIIA.				
DR	InterPro; IPR001996; PTS_EIIB.				
DR	InterPro; IPR003352; PTS_EIIC.				
DR	InterPro; IPR000504; RRM.				
DR	Pfam; PF00358; PTS_EIIA.1; 1.				
DR	Pfam; PF00367; PTS_EIIB.1.				
DR	Pfam; PF02378; PTS_EIIC.1.				
DR	PROSITE; PS002243; PTS_EIIA.1.				
DR	PROSITE; PS00371; PTS_EIIA.1; UNKNOWN_1.				



100

DA EMBL; AB00/002; AA00/001; -



DR InterPro: IPR001127; PTS\_EIIA.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR Pfam: PF00358; PTS\_EIIA\_1; 1.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD001476; PTS\_EIIB; 1.  
DR ProDom: PD002243; PTS\_EIIA; 1.  
DR ProSITE: PS00371; PTS\_EIIA\_1; UNKNOWN\_1.  
DR ProSITE: PS01035; PTS\_EIIB\_CYS; 1.  
KW Complete proteome.  
SQ SEQUENCE 628 AA: 67460 MW: 7F715A1D2B9DE19A CRC64;

Query Match 25.9%; Score 612; DB 16; Length 628;  
Best Local Similarity 34.1%; Pred. No. 2.4e-30;  
Matches 154; Conservative 72; Mismatches 179; Indels 46; Gaps 11;  
QY 2 AMVFPSLVNGYDVATMAAGEMPMSLF-----GLDVAQAGYQGTVPVLV 47  
Db 180 SLVYPTL-----SPLMTGK-PLYTLFQGTIFASPVVYTFGLGIPVILMSYSTVPIIL 231  
QY 48 VSWILATIEKFLHKLKRGTADELIITVLLLTGTLTFTTAIGPAMRWGDVLAHGLQGLY 107  
Db 232 ASYVGKVEKAKAIPDVVKTFVPECTLLVMVPLSLIVGPISTWAGKLLGAGTLAIY 291  
QY 108 DFGCPVGGILLGLVYSPVITGLHQSPPI---ELEFNQGGSFIFATASMANIAQCAAC 164  
Db 292 NLSPILAGLFIGAFWQVVFVIFGLHNGLVPIAMNNLSVLHYDP--ILAGTIGASFAQTGVV 349  
QY 165 LAVFFLAKSEKLGKLAGAGSVSAVLGITTEPAIFGVNLRNRWPFFIGTAAIGGALIALF 224  
Db 350 LAILIKTKNVKLGKIALPAFISGIFGVTEPAIYGVTLPRKKPFIISCICAAIGGGITGFM 409  
QY 225 NIKAVAGAAGFLGVVS-IDAPDMVMFLCAVVTFEFAFGAAIAYGLYLVRNRGSDIPDA 283  
Db 410 GTKLMMMGGLGIFAIPISYIGAKMDRGFGYGVMSVVISF----VVGFLIMEFAGPKDEE- 464  
QY 284 TRAPVPAGTTKAAEAPAEFSDNSTIIQAPLTGEATALSVDAMPASGLSGVAIVPT 343  
Db 465 ----VKQETTKKK-----NELVKQETLV-SPLKGIKTLSEVKDEAFSTGSLGKGIAPPE 515  
QY 344 KGLVSPVSGKIWVAPPSGHAFVRKAEKDSNVDLMHIGFTVNLNGTHFNPLAKQGD 403  
Db 516 EGKLVSPVGVGLATLEPTGHAVGIIS-----DKGTELLIHVGMTVOLEKGYFTILKQGD 571  
QY 404 EVKAGELICEFDIDAIAKAAGYEVTTPIVVSN 434  
Db 572 HVKAGDTILEFDIPKIKKAGYTLTTPVVVTN 602

Search completed: September 25, 2002, 06:07:18  
Job time: 504 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:02:04 ; Search time 1711.96 Seconds  
(without alignments)  
12038.745 Million cell updates/sec

Title: US-09-604-231-1

Perfect score: 1527

Sequence: 1 ctcattgcatctgcgcgtt.....gttgaaaccttgagtgctcg 1527

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	48.8	3.2	680	9	AL506262
C 2	45.8	3.0	473	10	T03017
C 3	45	2.9	605	12	BH175021
C 4	41.8	2.7	681	12	CNS02EOD
C 5	41.6	2.7	895	12	CNS0071A
C 6	41.4	2.7	477	10	BE997958
C 7	41.4	2.7	559	9	AW574083
C 8	41.4	2.7	601	10	BG580528
C 9	41.4	2.7	627	10	BE998413
C 10	40.8	2.7	884	12	CNS03CVD
C 11	40.8	2.7	983	12	CNS03CVD
C 12	40.8	2.7	1101	12	CNS00LXJ
C 13	40.2	2.6	970	12	CNS03H6V
C 14	39.8	2.6	606	10	BE997957
C 15	39.2	2.6	532	10	BE997957
C 16	39.2	2.6	993	12	BE997957
C 17	38.8	2.5	539	9	AI070146

18 38.8 2.5 714 9 AW583970  
C 19 38.6 2.5 401 9 AI436816  
C 20 38.6 2.5 1204 10 BG391833  
C 21 38.4 2.5 560 9 BB643902  
C 22 38.4 2.5 1101 12 CNS0021D  
C 23 38.2 2.5 500 10 BF727921  
C 24 38.2 2.5 525 12 CNS025EN  
C 25 38.2 2.5 531 9 AW573719  
C 26 38.2 2.5 621 10 BE998387  
C 27 38.2 2.5 652 9 AI982977  
C 28 38.2 2.5 687 9 AW980716  
C 29 38.2 2.5 703 9 AW980735  
C 30 38.2 2.5 713 10 BG581619  
C 31 38.2 2.5 737 10 BG581892  
C 32 38.2 2.5 789 10 BG582423  
C 33 38.2 2.5 795 10 BG583630  
C 34 37.8 2.5 654 9 AI812147  
C 35 37.6 2.5 726 10 BM014664  
C 36 37.6 2.5 752 10 BM423398  
C 37 37.6 2.5 780 10 BI118076  
C 38 37.6 2.5 1101 12 CNS00PXE  
C 39 37.4 2.4 460 9 AI318021  
C 40 37.4 2.4 552 10 BE721200  
C 41 37.4 2.4 741 10 BI856144  
C 42 37.4 2.4 904 12 CNS03HA0  
C 43 37.2 2.4 421 9 AW173566  
C 44 37.2 2.4 561 9 AI917973  
C 45 37.2 2.4 594 9 AW440291

#### ALIGNMENTS

RESULT 1

AL506262/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AW583970 la07h02.x  
AI436816 SMOV3MCAM  
BG391833 602418142  
BB643902 BB643902  
AL097099 Drosophil  
BF727921 1000055A0  
AL181976 Tetraodon  
AW573719 EST316310  
BE998387 EST430110  
AI982977 wt46c03.x  
AW980716 EST391869  
AW980735 EST391888  
BG581619 EST483354  
BG581892 EST483628  
BG582423 EST484167  
BG583630 EST485382  
AI812147 605087H03  
BM014664 603640A93  
BM423398 AGENCOURT  
BI118076 602867346  
AL071370 Drosophil  
AI318021 SMOVAFAP  
BE721200 188296 MA  
BI856144 603382851  
AL244017 Tetraodon  
AW173566 x108h03.x  
AI917973 tz16d08.x  
AW440291 hb89a03.x

AL506262 680 bp mRNA linear EST 04-JAN-2001  
AL506262 Hordeum vulgare Barke developing caryopsis (3'-15.DAP)  
Hordeum vulgare cDNA clone HY021247 5', mRNA sequence.

AL506262

EST

barley.

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticaceae; Hordeum.

1 (bases 1 to 680)

Michalek W., Weschke W., Pleissner K.-P. and Graner, A.

EST sequencing and analysis in barley

Unpublished (2000)

Contact: Michalek W

Institute for Plant Genetics and Crop Plant Research

Corrensstr.3, D-06466 Gatersleben, Germany

Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

Seq primer: T3 primer for 5'end.

Location/Qualifiers

1..680

/organism="Hordeum vulgare"

/cultivar="Barke"

/db\_xref="taxon:4513"

/clone="HY021247"

/clone\_lib="Hordeum vulgare Barke developing caryopsis

(3'-15.DAP)"

/lab\_host="XLOLR"

/note="vector: plasmid pBK-CMV; Site\_1: EcoRI; Site\_2:

XhoI; mRNA was made from developing caryopsis (3'-15.DAP)

of spring barley variety 'Barke', a high quality malting

variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI

(3'-end of cDNA). NOTE: Due to a cloning artefact caused

by the kit, in most cases the EcoRI site is NOT present,

as well as the EcoRI adapter. Average insert size is 1 kb

Sequence trimming: Vector sequences and sequence ends were

**Figure 1**

```
/strain="multiple"
/db_xref="taxon:1639"
/clone="63-43"
/clone_lib="Listeria monocytogenes pUC18 library"
/note="vector: pUC18; Shotgun library prepared by Amplicon Express (Pullman, WA)"
BASE COUNT      154 a 110 c 125 g 216 t
ORIGIN

Query Match      2.9%; Score 45; DB 12; Length 605;
Best Local Similarity 46.8%; Pred. No. 0.39;
Matches 141; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 210 gttaccagggaccgtgtcttcctgctggtggtgtttcttctggtattctggtcagcagcatcgaga 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GTATATCTTCTCCGTTATTCGAATATATCTTGGCAAGTTTGGTTTATATCTATTTAGAAA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 270 agttctgcacacgagcactcaaggcactgcagacttctgtatcactccagtcgctgacgt 329
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GATTTCCTTAATAGTAAATTCATGAGCGACGCAAGACCTTTTAAACGCCGATGATTGCC 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 330 tgtctcacccggtattcttaccattcattcgtccatggccagcaatgcgctgggtggcg 389
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TGATGCTTATCGTACCGCTTACATCTTACGCTTTGGACCTCTGGTACGTTATTAGTC 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 atgtctgcacacggtctacaggacatttatgatttcggtgtccagtcggcggtctgc 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AAGGTTTACGTCAGCGGTATACATTTATTAACCTTGAGCCCAATTTGAGCGGGTGCTT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 450 tcttcggtctgtctactcaccacatcgcatcactggtctgcaccagctctcccgcccaa 509
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TTATCGGGTGGTCTCGCAAGATATTAGTTATTTTGGGATTCATTGGGGCTTTGTGCCAA 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 510 t 510
Db 481 T 481

RESULT 4
CNS02EOD
LOCUS      681 bp DNA linear GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 262H14 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION ALI93390
VERSION   ALI93390.1 GI:7832096
KEYWORDS  GSS; genome survey sequence.
SOURCE    Tetraodon nigroviridis.
ORGANISM  Tetraodon nigroviridis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
           Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 681)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
           Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
           Weissenbach,J.
TITLE     Characterization and repeat analysis of the compact genome of the
           freshwater pufferfish Tetraodon nigroviridis
JOURNAL   Unpublished
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
TITLE     Human gene number estimate provided by genome wide analysis using
           Tetraodon nigroviridis DNA sequence
JOURNAL   Unpublished
AUTHORS   3 (bases 1 to 681)
           Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT   This sequence is a single read and was generated as part of a large
```

```
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES             source
Location/Qualifiers
1..681
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="262H14"
/clone_lib="G"
/note="Genoscope sequence ID : COAG262DD07LPI-end : T7"
BASE COUNT      74 a 105 c 170 g 282 t 50 others
ORIGIN

Query Match      2.7%; Score 41.8; DB 12; Length 681;
Best Local Similarity 40.9%; Pred. No. 3;
Matches 139; Conservative 18; Mismatches 183; Indels 0; Gaps 0;

QY 632 ggccttcaggtgcttcaggtgtctccgctgtctcttggtattacgagcctgcgactctc 691
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 GCGGTGTYGTGTTTGTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 692 ggtgtgaacctgcctgcctgcctggcgtctcttcacgtatcggtaccgacgtatcgt 751
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 752 ggcgtttgatgcactcttataatcaaggcagtgctgctggcgctcaggtttctctg 811
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 GTNGTGTGTTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 812 ggtgtttcttatgatctccagatatggtcatgtctctctgtgtgtgacattgttacc 871
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 GTGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 872 ttcttcacgtcattcggcgacgagattgcttatggccttacttggctgcgcgcaacgcg 931
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 GCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 agcattgatccagatgaaccgctgctccagtcgctcag 971
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
CNS0071A
LOCUS      895 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
           BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
           fly), genomic survey sequence.
ACCESSION AL066286
VERSION   AL066286.1 GI:4945153
KEYWORDS  GSS.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 895)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
           collaboration with the Berkeley Drosophila Genome project (BDGP).
           The BDGP is constructing a physical map of the Drosophila
           melanogaster genome using these BACs. For further information
           please see http://www.fruitfly.org The BDGP Drosophila
           melanogaster BAC library was prepared by Kazutoyo Osoegawa and
           Aaron Mammosier in Pieter de Jong's laboratory in the Department of
           Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
           NY. The library is named RPCI-98 and was constructed by partial
           EcoRI digestion of Drosophila DNA provided by the BDGP from the
```









[illegible]

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 970)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 970)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 970)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
source  
1..970  
    /organism="Tetraodon nigroviridis"  
    /db\_xref="taxon:99883"  
    /clone\_xref="026011"  
    /clone\_lib="G"  
    /note="Genoscope sequence ID : COBG026AH06LP1-end : T7"  
BASE COUNT     311 a   298 c   194 g   152 t       15 others

Query Match                  2.6%; Score 40.2; DB 12; Length 970;  
Best Local Similarity        45.9%; Pred.No. 9.8;  
Matches 130; Conservative    2; Mismatches 151; Indels    0; Gaps    0;

QY 637 tcaggttcagggtgcctcgcctgttcttgattacaggacctgcgatcttcgggtg 696  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 709 TGCTGTGTTGTTGCTGCTGCTGTTGTTGTTGTTGCTGCTGCTGTTGTTGCTGT 650  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 697 gaaccttcgcctgcctgcccgtcttcctcatcggtatcggtaccgcagctatcggtgccgc 756  
+  
Db 649 TGTTGCTGCTGCCTGCTGCTGTTGTTGTTGCTGCTGCTGTTGATGATGCTGCTGTTGTTGT 590  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 757 ttgattgcactcttaataatcaaggcaggttcggtggcgctgcaggtttcttcgggtgt 816  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 589 TGCTGTGCTGCTGTTGTTGTTGCTGCTGCTGTTGTTGCTGCTGCTGTTGTTGCTGTC 530  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 817 tgtttotattgatgctccagatatggtcattcttggtgtgtgcaggtttctaactcttt 876  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 529 TGCTGTTGTTGTTGCTGCTGCTGTTGTTGCTGCTGTTGTTGCTGCTGTTGTTGCTGTTGT 470  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 877 catcgattcggcgacgagtcttatggccttacttgttt 919  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 469 TGTTGCTGCTGCTGCTGTTGTTGCTGCTGCTGTTGTTGCTGTTGTTGTTGTTGTT 427  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14  
BE997957  
LOCUS BE997957 606 bp mRNA linear EST 06-OCT-2000  
DEFINITION EST429680 GVSN Medicago truncatula cdna clone pgVSN-8B9, mRNA  
sequence.  
ACCESSION BE997957  
VERSION BE997957.1 GI:10698233  
KEYWORDS EST,  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae Medicago.

1 (bases 1 to 606)

Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.

ESTs from senescent nodules of *Medicago truncatula*

Unpublished (2000)

Contact: Carroll P. Vance

Department of Agronomy and Plant Genetics

University of Minnesota

411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA

Tel: 612 625 5715

Fax: 651-649-5058

Email: vance004@maroon.tc.umn.edu

University of Minnesota name: M271935e TIGR sequence name: MTKAK050K More information is available at: http://chrysis.tamu.edu/medicago

Seq primer: SKnod (CTA gAA CTA gTg gAT CC).

Location/Qualifiers

1..606

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db\_xref="taxon:3880"

/clone\_lib="pGVSN-8B9"

/clone\_lib="GVSN"

/tissue\_type="senescent root nodules"

/dev\_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"

/lab\_host="E. coli strain SOLR"

/note="Vector: pBluescript SK +/-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 173 a 122 c 127 g 184 t

ORIGIN

Query Match 2.6%; Score 39.8; DB 10; Length 606;

Best Local Similarity 49.8%; Pred. No. 9.9;

Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1246 cattgggttcgacacagtaaaacctcaacggcgacgactttaaccgctgagaagcagg 1305

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 317 CATTGACTTCCCAAGAAGAGAGTTCTCATAGCTTTCTTAAGCCTTAACGGATCATAA 376

QY 1306 cgatgaagtcaaacgaggggagctgctgtgtgaattcgattgatgcattaaagctgc 1365

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 377 TGCTTTTGTCACGTGCTCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAAGCGAC 436

QY 1366 aggttatgggttaaacacgcgcgattgtgttctgaattacaagaacccgacctgtaaa 1425

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 437 CAATGAAGAAGTAGCGGCACCTTTTCGTGAGGCTGATTCTGTAGTCATGCATCATTAAG 496

QY 1426 cacttcacggtttgggcgaattg 1448

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 497 CTTTAAAGAGTTTGTCAAGCTTG 519

RESULT 15

BI856505/c

LOCUS

DEFINITION

60385596F1 NIH\_MCC\_87 Homo sapiens cDNA clone IMAGE:5394670 5', mRNA sequence.

ACCESSION

BI856505

VERSION

BI856505.1 GI:15997252

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 532)

TITLE

NIH-MGC <http://mgc.nci.nih.gov/>.

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM12005 row: 0 column: 23  
High quality sequence stop: 522.

FEATURES

Location/Qualifiers

source

1..532  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5394670"  
/clone\_lib="NIH\_MGC\_87"  
/tissue\_type="mammary adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: breast; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT

185 a 122 c 134 g 91 t

ORIGIN

Query Match

2.6%; Score 39.2; DB 10; Length 532;

Best Local Similarity

47.2%; Pred. No. 13;

Matches 119; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY

757 ttgatgtcactctttaataatcaaggcagttgcgttggcgctgcaggtttcttgggtgt 816

Db

472 TATGATTGCCACTTATGCTTTCCTGCAGATGCTCTGGAGCAGCTGCTTTCGGGCTGG 413

QY

817 tgtttctattgatgtccagagatatggtcattcttgggtgtgtgcagttgtaccttctt 876

Db

412 AGCTTCTTGACCTTCTGAGCTTCTGAGCAGCGCTGCTTCTTGGCCTTGGCTTCTG 353

QY

877 catcgcatteggcgagcgattgttattgaccttacttggttcgcgcaacgagcat 936

Db

352 GGCAGGNACCTTCTGGCTGGAGCCCTTTTACTCGGGCGGTGATCTTTTGTGGAC 293

QY

937 tgatccagatgcacacgctgtcctcagtcgaggaacccaagccgaagcagaagc 996

Db

292 TTTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 233

QY

997 acccgagaatt 1008

Db

232 AGCAGCAGCAGT 221

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:07:24 ; Search time 54.45 Seconds  
(without alignments)  
6888.565 Million cell updates/sec

Title: US-09-604-231-1  
Perfect score: 1527  
Sequence: 1 ctcagcgatcgccgtt.....gtgaaaccttgagtgctg 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length  | ID                   | Description        |
|------------|-------|-------------|---------|----------------------|--------------------|
| C 1        | 109.6 | 7.2         | 3615    | 1 US-08-920-812-17   | Sequence 17, Appl  |
| C 2        | 109.6 | 7.2         | 3615    | 1 US-08-920-827-17   | Sequence 17, Appl  |
| C 3        | 109.6 | 7.2         | 3615    | 1 US-08-921-177-17   | Sequence 17, Appl  |
| C 4        | 109.6 | 7.2         | 3615    | 1 US-08-362-577C-17  | Sequence 17, Appl  |
| C 5        | 109.6 | 7.2         | 3615    | 2 US-08-920-828-17   | Sequence 17, Appl  |
| C 6        | 97.8  | 6.4         | 465     | 2 US-08-920-828-17   | Sequence 3, Appl   |
| C 7        | 71.6  | 4.7         | 357     | 2 US-08-673-190A-3   | Sequence 3, Appl   |
| C 8        | 43.2  | 2.8         | 7218    | 1 US-08-673-190A-6   | Sequence 6, Appl   |
| C 9        | 39.8  | 2.6         | 1314    | 4 US-08-232-463-14   | Sequence 14, Appl  |
| C 10       | 37.2  | 2.4         | 4403765 | 4 US-08-936-165A-153 | Sequence 153, Appl |
| C 11       | 34.6  | 2.3         | 46899   | 1 US-09-103-840A-2   | Sequence 2, Appl   |
| C 12       | 32.8  | 2.1         | 12912   | 1 US-08-471-119A-1   | Sequence 1, Appl   |
| C 13       | 32.8  | 2.1         | 14060   | 3 US-08-460-751-1    | Sequence 1, Appl   |
| C 14       | 32.6  | 2.1         | 477     | 4 US-08-658-136-4    | Sequence 4, Appl   |
| C 15       | 32.6  | 2.1         | 80161   | 3 US-09-135-994-1    | Sequence 1, Appl   |
| C 16       | 32.6  | 2.1         | 80161   | 3 US-09-036-987A-1   | Sequence 1, Appl   |
| C 17       | 32.4  | 2.1         | 292     | 4 US-09-370-700-1    | Sequence 1, Appl   |
| C 18       | 32.2  | 2.1         | 7218    | 4 US-09-117-121-29   | Sequence 29, Appl  |
| C 19       | 32    | 2.1         | 1008    | 4 US-08-232-463-14   | Sequence 14, Appl  |
| C 20       | 32    | 2.1         | 1901    | 4 US-09-198-955A-9   | Sequence 9, Appl   |
| C 21       | 32    | 2.1         | 1901    | 3 US-08-153-848-43   | Sequence 43, Appl  |
| C 22       | 32    | 2.1         | 1901    | 4 US-09-299-843A-43  | Sequence 43, Appl  |
| C 23       | 32    | 2.1         | 1901    | 4 US-09-088-337B-43  | Sequence 43, Appl  |
| C 24       | 32    | 2.1         | 2232    | 5 PCT-US93-11153-43  | Sequence 43, Appl  |
| C 25       | 32    | 2.1         | 2453    | 5 PCT-US95-07180-1   | Sequence 12, Appl  |
| C 26       | 32    | 2.1         | 3120    | 1 US-08-241-766-12   | Sequence 1, Appl   |
| C 27       | 32    | 2.1         | 3120    | 1 US-08-491-146-1    | Sequence 1, Appl   |
| C 28       | 32    | 2.1         | 3120    | 1 US-08-241-766-11   | Sequence 11, Appl  |

|      |      |     |       |                     |                   |
|------|------|-----|-------|---------------------|-------------------|
| 28   | 32   | 2.1 | 3120  | 1 US-08-234-011-1   | Sequence 1, Appl  |
| 29   | 32   | 2.1 | 3120  | 2 US-08-701-062A-1  | Sequence 1, Appl  |
| 30   | 32   | 2.1 | 3533  | 4 US-08-985-916-15  | Sequence 15, Appl |
| C 31 | 32   | 2.1 | 5163  | 3 US-08-700-651-1   | Sequence 1, Appl  |
| C 32 | 32   | 2.1 | 5163  | 3 US-08-928-361B-4  | Sequence 4, Appl  |
| C 33 | 32   | 2.1 | 5318  | 3 US-08-700-651-2   | Sequence 2, Appl  |
| C 34 | 32   | 2.1 | 5318  | 3 US-08-928-361B-3  | Sequence 3, Appl  |
| C 35 | 32   | 2.1 | 10348 | 2 US-08-457-273B-41 | Sequence 41, Appl |
| 36   | 32   | 2.1 | 10348 | 3 US-08-556-419-13  | Sequence 13, Appl |
| 37   | 32   | 2.1 | 10348 | 4 US-09-041-886-14  | Sequence 14, Appl |
| 38   | 32   | 2.1 | 10366 | 1 US-08-246-982A-5  | Sequence 5, Appl  |
| 39   | 32   | 2.1 | 10366 | 1 US-08-453-265-5   | Sequence 5, Appl  |
| C 40 | 31.8 | 2.1 | 1227  | 3 US-09-074-912-3   | Sequence 3, Appl  |
| C 41 | 31.8 | 2.1 | 1227  | 4 US-09-290-136-3   | Sequence 3, Appl  |
| C 42 | 31.6 | 2.1 | 2336  | 4 US-09-228-986-10  | Sequence 10, Appl |
| C 43 | 31.6 | 2.1 | 5511  | 3 US-08-928-361B-2  | Sequence 2, Appl  |
| C 44 | 31.6 | 2.1 | 7334  | 3 US-08-928-361B-1  | Sequence 1, Appl  |
| C 45 | 31.6 | 2.1 | 7655  | 1 US-08-619-554-1   | Sequence 1, Appl  |

ALIGNMENTS

RESULT 1  
US-08-920-812-17/C  
; Sequence 17, Application US/08920812  
; Patent No. 5763188  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/920,812  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-24  
; US-08-920-812-17

Query Match 7.2%; Score 109.6; DB 1; Length 3615;  
Best Local Similarity 46.3%; Pred. No. 2.3e-23;  
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;

Qy 312 tcaactcagtgctgacggtgtctgctcaccggtatcttaccattcatcgcattggccag 371  
Db 2495 TCACACCAATGCTATGCTGATGGTTATACACACCGTCACTTCTGCTGGTGGGGCGC 2436  
Qy 372 caatcgctgggtggcgatgctgctggcacagcgtctcaccggagctttatgattcggtg 431  
Db 2435 TATCAACCTGGATAGCGAACTGATGGCGGGTATCTCTGGCTTTATCAGGGGTTTC 2376  
Qy 432 gtccagtcggcggtgctctcttcgtctgtgtctactaccacatgctactggtctgctgc 491  
Db 2375 CTGCATTTGGCGGCGCGTAAATGGCGGCTCTTGCAAAATCTGCTGATCTGGACTGC 2316  
Qy 492 accagtcctcccccacattgagctggagctgtttaaccagggtggatccttccttcg 551  
Db 2315 ACTGGGGCTGGTCCGGTGTGTAACAATACTTCAACCGTGTGGCTGACGACCACTGA 2256  
Qy 552 caacgcatctatg---gctaatacgcgccagggtgcgcatgtttggcagtgcttccc 608  
Db 2255 TCCCGCTTAATGCCCCCATTAATGGCGCAGGTGCGGGCGGCTCGCGCTCTCTCT 2196  
Qy 609 tggcgaaagtgaaagtcgaaggcccttgagcgtgcttcagtgcttcacgctgtctcttgg 668  
Db 2195 GCGAACCGGATCGCGAGAAAAAGTGGTGGGGATACAGCGCGTTGACGAGTCTGTTG 2136  
Qy 669 gtattcagagcctgcgacttcggtgtaaccttcgctgctgctgctggcggttcttcacg 728  
Db 2135 GTATCACCAGAACCGGATATGTCGTCACCTGCGCGTAAGTACCCCTTTGTTATCG 2076  
Qy 729 gtacgggtaccgcagctacgtggtggcgtttgattgactctttaaatacagcaggtg 788  
Db 2075 CTGTATCAGTGGGGCTTTGGGGGCCACCATTAATTTGGTACGCGCAACGAAGTCTACT 2016  
Qy 789 cgttggcgctgcagggttcttgggtgtgtttctctattgactcaccagatggtcatgt 848  
Db 2015 CTTTTGGTTGCAAGTATTTTCACTTCATGCAAAACCATCCGTCACAGGGAATTGAT 1956  
Qy 849 tcttggtgtgacagtggttacccttcttcacgacttcgacttcgagcagcagattcttatgcc 908  
Db 1955 TCACCGTCTGGCCAGCGCTTATTTGGCGGTGTCATTTGCCATCGGTGGCCATTTGTCGTA 1896  
Qy 909 ttacttggttcgcgcacgcagcagcttgatccagatgcaacgcgtctccagtgccctg 968  
Db 1895 CGGTGATGCTTTCATTCATCACCCTAAACCTCAGCCAGCGCAGGTGCCCC----- 1842  
Qy 969 caggaaacacaaagcgaagcagaagcaccgcagaaatttcaaacgattccaccatca 1028  
Db 1843 -----GAAGAGAAAAACACAGAGTTATTACACCACCTGAGCAGGGCGGTA 1798  
Qy 1029 tccaggcacctttacgcgtgaagctattgcactgagcagcagcagcagcagcagctgtttg 1088  
Db 1797 TCTGTTACCATGACGGGAGAGATTGTGCGCTCATTTCACTGCTGCTGATACCACTTTG 1738  
Qy 1089 ccagcgaaagcttgctcgccggtgctccatcgtcccaacaaagggcagttagttctc 1148  
Db 1737 CCAGTGGCGCTGTGGTAAAGTATGGCATTCCTCCCTCGCTGGTGGTCAAGTGGTCTC 1678  
Qy 1149 cggtagtggaagattggtggcattcccatctgtgccaatgcttcttcgagcttcgaccca 1208  
Db 1677 CGGTTGGGGTCAATGCTTCTGTTGTTGCGCCACATTACA-----CGCCATTGGCATTTG 1624  
Qy 1209 aggtcagagtggttccaatgatacttctgacatctgacattggttttcgacagataaac 1268  
Db 1623 AGTCAGATGATGG-----TGTGGAGATCTCTGATTCATGTCGGTATCGGACACCGTAAAC 1570  
Qy 1269 tcaacgcgcagcactttaaccgcgtgaagaagcagggcgatgaagtcaaacagcagggagc 1328  
Db 1569 TGGACGGCAAAATCTTTTCCGCTCAGCTCAACGTGGGTGACAAAGTCAATACAGCGCATC 1510

Qy 1329 tgcgtgtgaattgatattgatgccattaaggctgcagggttatgagtgtaaacacgcgca 1388  
Db 1509 GCGTGAATTTCTTTGATATCCTGCTATTCGAGGCGGGATTTGATCTGACGACGCGG 1450  
Qy 1389 ttgtgttttcaattacagaacacccggacactgtaaacacattacggttttggcgaattg 1448  
Db 1449 TATTAACTAGTAATGACGATGATTTACGGAGCTATTACCCACGCGCAGCGCAGATAA 1390  
Qy 1449 aagcgggagcacaacctgctcaacgtc 1474  
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

## RESULT 2

US-08-920-827-17/c  
; Sequence 17, Application US/08920827  
; Patent No. 5770375  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,827  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical isolate EC-24  
; US-08-920-827-17

Query Match 7.2%; Score 109.6; DB 1; Length 3615;  
Best Local Similarity 46.3%; Pred. No. 2.3e-23;  
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;

Qy 312 tcaactcagtgacggtgtgctcaccggatcttaccattcatcgcattggccag 371  
Db 2495 TCACACCAATGCTATGCTGATGGTTATACACACCGTCACTTCTGCTGGTGGGGCGC 2436  
Qy 372 caatcgctgggtggcgatgctgctgcaacacggtctacagggactttatgattcggtg 431

|    |      |   |      |
|----|------|---|------|
| Db | 2435 | TATCAACCTGGATTAACGAACTGATTGGCGCGGTATCTCTGCTGCTTATACAGCGGTTCT    | 2376 |
| Qy | 432  | gtccagtcgcggtctgctcttcggtctctggtctactaccaatcgctcatcaactggtctgc  | 491  |
| Db | 2375 | CTGCATTTGGGGCGGGTAATGGCGGCTTCTGGCAAACTCTTCGTCAATGTTTCGGACTGC    | 2316 |
| Qy | 492  | accagtcctcccgccaaatgagctggagctgtttaaccagggtggatccttcattcttcg    | 551  |
| Db | 2315 | ACTGGGGCCTGGTGGCGGTGTATCAATAACTTCACCGTCTGGGCTACGACACCATGA       | 2256 |
| Qy | 552  | caacgcatctatg---gctaataatcgccagggtgaggaatggttggcagtggtcttc      | 608  |
| Db | 2355 | TCCCGCTGTTAATGCCCGCATATATGGCGCAGGTGCGGGCGGCGCTCGGGGTCTTCCTCT    | 2196 |
| Qy | 609  | tggcgaagagtgaaaaagctcaaggccctgcaagtgcttcaggtgtctccgctgtctctg    | 668  |
| Db | 2195 | GCGAACCGCATGCGCAGAAAAAGTGTGTGGCGGGATACGCGCGTTCACGAGTCTGTTTC     | 2136 |
| Qy | 669  | gtatacggagcctcgatctcttggtgtgaaccttcgctcgctgcgtgccttcttcacg      | 728  |
| Db | 2135 | GTATCACGAAACGACGGTATATGGCGTCAACCTGCGCGTAAGTACCCTTTGTTATCG       | 2076 |
| Qy | 729  | gtatcggtaaccgagctatcggtagcgctttgattggaactctttaataatacaaggcagttg | 788  |
| Db | 2075 | CTGTATACGTGGGGCTTTGGGGGCCACCATTTTGGTACGCGCAACGAAAGTCTACT        | 2016 |
| Qy | 789  | cgttgggcctgcaggttcttgggtgtgttcttatgatgtccagatatggtcatgt         | 848  |
| Db | 2015 | CCTTTGGTTGCCAAGTATTTTTCACCTTCATGCAAAACCATCCGTCACGGGAATTGATT     | 1956 |
| Qy | 849  | tcttgggtgtgcagttgttacctcttcacgatactcggcgagcagattgcttatggcc      | 908  |
| Db | 1955 | TCACCTCTGGGCCAGCGTTATTGGCGGTGTCATTGGCCATCGGTGCGCATTTGTCGGTA     | 1896 |
| Qy | 909  | tttaactgtttgcgcgaacggcagcattgatccagatgcgaacgcgtgctccagtgctg     | 968  |
| Db | 1895 | CGGTGATGCTTCATTTTCACCGGTAAACGTACAGCCAGCGCAGGGTGCCTC-----        | 1842 |
| Qy | 969  | caggaaacgacaaagccgaagcagacgacccgcagaatttccaaacgattccaccatca     | 1028 |
| Db | 1843 | -----GAAGAGAAACACACAGAGGTTATTACACCACTGACGAGGCGGTA               | 1798 |
| Qy | 1029 | tccagcaaccttgaccggtgaagctattgcactgagcagcgctcagcagtgccatgtttg    | 1088 |
| Db | 1797 | TCGTTCACCGATGACGGGAGAGATTGTGCGCTCATTCACGTCGCTGATACACAGTTTG      | 1738 |
| Qy | 1089 | ccagcgaagaacttgctcggcgcttgccatcgtcccaaccaaagggcagtgatttctc      | 1148 |
| Db | 1737 | CCAGTGGCCTGTGGTAAAGTATTTGGCATCTCGCCTCGTTCGTTGCTGAAGTCGGTTCTC    | 1678 |
| Qy | 1149 | cggtagtggaagatgttggcgcatcccatctggccatgctttcgcagttcgaccca        | 1208 |
| Db | 1677 | CGGTTCGGGGTCCAATGTCTTGTTGTCGCCACATTACA-----GCCATTGGCATTTG       | 1624 |
| Qy | 1209 | aggtcagatggttccaatgtggatatcttgatgcacattggttttcgacacagtaaac      | 1268 |
| Db | 1623 | AGTCAGATGATG-----TGTGGAGATCCTGATTCTATGTCGGTATCGACACGCTAAAC      | 1570 |
| Qy | 1269 | tcaacggcgcgcactttaaccgcgtgaagaagcagggcgatgaagtcaaacgagggagc     | 1328 |
| Db | 1569 | TGACCGGCAAAATCTTTTCCGCTCACAGTCAACGTCGGGTGACAAGGTCAATACAGCCGATC  | 1510 |
| Qy | 1329 | tgcctgtgaattcgatattgatcattaaagctgcaggttatgaagtaaacacgccga       | 1388 |
| Db | 1509 | GGCTGATTTCTTTTGATATCCCTGCTATTTCGGGAGGCGGATTTGATCTGACAGCGCGG     | 1450 |
| Qy | 1389 | tgtgttttcgaattacaagaacacggacacctgtaaacacttacggtttgggggaattg     | 1448 |
| Db | 1449 | TATTATCAGTAATACGATGATTTTACGGAGCTATTACCCTACGGCACGGCGCAGATAA      | 1390 |
| Qy | 1449 | aagcgggaccaaactgctcaacgtc                                       | 1474 |

Db 1389 GCGAGGTGAACCGCTGTTCATC 1364

RESULT 3

US-08-921-177-17/c

; Sequence 17, Application US/08921177

; Patent No. 5798211

; GENERAL INFORMATION:

; APPLICANT: Ohno, Tsuneya

; APPLICANT: Matsuhisa, Akio

; APPLICANT: Uehara, Hirotsugu

; APPLICANT: Ega, Soji

; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/921,177

; FILING DATE: 29-AUG-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/362,577

; FILING DATE: 27-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Rin-Laures, Li-Hsien

; REGISTRATION NUMBER: 33,547

; REFERENCE/DOCKET NUMBER: 19036/32420

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/474-6300

; TELEFAX: 312/474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3615 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli

; STRAIN: Clinical Isolate EC-24

US-08-921-177-17

| Query Match           | 7.2%;  | Score 109.6;       | DB 1;      | Length 3615; |
|-----------------------|--|--------------------|------------|--------------|
| Best Local Similarity | 46.3%;   | Pred. No. 2.3e-23; |            |              |
| Matches 540;          | Conservative 0;  | Mismatches 589;    | Indels 37; | Gaps         |
| Qy 312                | tcactccagtcgcagctgctgcacgcgattcttacattcatcgccattgcccag         | 371                |            |              |
|                       |  |                    |            |              |
| Db 2495               | TCACACGATTCCTATCTCTGATGGTTATCACACCGTCACCTTCTCTGCTGGTGGGGCCG    | 2436               |            |              |
| Qy 372                | caatgcgcctggggcgatgctggcacacgcgtctacaggggactttatgatttcgggtg    | 431                |            |              |
|                       |  |                    |            |              |
| Db 2435               | TATCAACCTGGATAAGCGAAGCTGATTGCCGCGGTTATCTCTGGCTTTATCAGCGCGTTC   | 2376               |            |              |
| Qy 432                | gtccagtcggcgctgcctctcgctcgtctggtctactaccaatgctcatcgtgtctgc     | 491                |            |              |
|                       |  |                    |            |              |
| Db 2375               | CTGCATTTGGCGGCGGGTAATGGGGGGCTTCTGCACAACTCTCTCATGTTCGGAGTCGC    | 2316               |            |              |
| Qy 492                | accagtcctcccgcccaattgagctggagctgtttaaccaggggtggatccttcattctcg  | 551                |            |              |
| Db 2315               | ACTGGGGCCCTGGTGGCGGTTGTATCAATTAACCTACACGCTGTGGGCTTACCACACCATGA | 2256               |            |              |

```
Qy 552 caacggcatctatg---gctaataatgcccccagggtgcggcatgttttggcagtggtttcttc 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 TCCGCGCTGTTAAATGCCCGCCATTATATGCGAGTGCAGTTCGGGGCGGCTGCGCTCTCTCT 2196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 609 tggcgaagatgaaactcaaggcccttcagggtgttcagggtgttcagggtgttcagggtgttc 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2195 GCGAACCGATGCGCAGAAAAAGTGTGGCGGATGAGCGGGCTGACGAGTCTGTGTTG 2136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 669 gtattacggagcctgcagatcttcgggtgtgaaccttcgggtgttcagggtgttcagggtgttc 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2135 GTATCACCGRACAGCGGTATATGGCGTCAACCTGCGCGGTAGTACCCCTTCTTATCG 2076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 729 gatacggatgcagctgcgggtgcgcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2075 CCTGTATCAGTGGGCTTGGGGCCACCATTATGCGTACGCGCAACCAACGAAAGTCTACT 2016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 789 cgttggcgctgcaggtttcttgggtgtgtttcttgcgttgcgttgcgttgcgttgcgttgcgt 848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 CCTTGTGTTGCCAAGTATTTTACCTTCATGCAAAACCATCCCGTCAACGGGAATTGATT 1956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 849 tcttgggtgtgcagttgttaccttcttcacgttcgcatcgcgcgcagcgttgcgttgcgttgc 908
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1955 TCACCGTCTGGCCAGCGTTATTTGGCGGTGTCTATGCGCATCGGTGGCGATTGTCGGTA 1896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 909 ttacttgggttcgcgcgaagcagcattgatccagatgcacccgctgcctccagtgctg 968
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1895 CGGTGATGCTTCAATTCATCACCGCTAAACGTACGCGCAGCGGTGCC----- 1842
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 969 caggaaagcaaaagcgaagcagaagcagccgcagaaatttcaaacgattccaccatca 1028
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1843 -----GAAGAGAAACACACAGAGTTATTACACACACCTGAGCAGGGCGGTA 1798
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1029 tccaggcacctttgacgggtgaagctattgcaactgagcagcgtcagcgtatgctatgtt 1088
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1797 TCTGTTTACCAGTACCGGAGAGATTGTCGCTCATTCAGTTCGCTGATACACGTTG 1738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1089 ccagcgaaagcttggctcggcggttccatgcgttcccaacaaagggcagttgattctc 1148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1737 CCAGTGGCGTGTGGTAAAGGTATTGCCATTTCGCTCGCTCGGTGGTGAAGTGGTCTC 1678
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1149 cggtagtgaaagattggtggcattcccatcctcctcctcctcctcctcctcctcctc 1208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1677 CGGTGGCGGTGCGAATGCTGCTGTTGTCGCCATTATCA-----GCGCATGGCGATTG 1624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1209 aggtcggaggtgttccaatgtgatatcttgatgcacattggtttcgacacagtaaac 1268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1623 AGTCAGATGATG-----TGTGGAGATCCTGATTCTATGTCGGTATCGACACCGTAAAC 1570
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1269 tcaacggcacgcaatttaacccgctgaagcagcagggcgatgaagtcaaacgaggggagc 1328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1569 TGGACGGCAAAATCTTTCCGCTCACGTCAACGTTGGTGACAAAGTCAATACAGCGCATC 1510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1329 tgctgtgattgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat 1388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 GCGTGATTTCTTTGATATCCCTGCTATTTCGGAGGCGGATTTGATCTCAGCAGCGCG 1450
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1389 ttgtgttgcgaattcaagaacacggcagcgtgtaaacacttacggtttgggggaaattg 1448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1449 TATTAATCAGTAATAGCGATGATTTTACGGACGCTATTACCCACGGCAGCGCGCAGATAA 1390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1449 aagcggagcgaacactgtcacaagtc 1474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 4

```
US-08-362-577C-17/c
; Sequence 17, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsunisa, Akio
```

```
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER TYPE: Floppy disk
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 3615 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORGANISM: Escherichia coli
STRAIN: Clinical Isolate EC-24
US-08-362-577C-17
```

```
Query Match 7.2%; Score 109.6; DB 1; Length 3615;
Best Local Similarity 46.3%; Pred. No. 2.3e-23;
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;
```

```
Qy 312 tcaactcagtgctgacgttgcgtcaccggtcccttaccatccatccatccatccatccatccatcc 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2495 TCACACCATTTGCTATGCTGATGTTATCATACCCGTCACCTTTCTGCTGGTGGGGCGC 2436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 372 caatgcgctgggtggcgatgtgctggcacacggtctcacaggggactttatgatttcggtg 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2435 TATCAACCTGGATAAGCGAACTGATTGCCGCGGTTATCTCTGGCTTTATCAGSCGGTTC 2376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 432 gtccagtcgcggtctgctcttccttccttccttccttccttccttccttccttccttccttc 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2375 CTGCATTTGGCGGCGCGGTAAATGGGCGGCTTCTGGCAAAATCTTCGTCATGTTCCGACTGC 2316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 492 accagtccttcgcccaattgagctggagctgttttaaccaggggtggatccttccttccttc 551
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2315 ACTGGGCGCTTGGTCCCGGTGTGATCAATACTTACCGTGTGGGTACGACACCAATGA 2256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 552 caacggcatctatg---gctaataatgcccccagggtgcggcatgttttggcagtggtttcttc 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 TCCGCGCTGTTAAATGCCCGCCATTATGCGCAGGTTCGGGGCGGCTGCGGCTCTCTCTCT 2196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 609 tggcgaagatgaaactcaaggcccttcagggtgttcagggtgttcagggtgttcagggtgttc 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2195 GCGAACCGATGCGCAGAAAAAGTGTGGCGGATGAGCGGGCTGACGAGTCTGTGTTG 2136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 669 gtattacggagcctgcagatcttcgggtgtgaaccttcgggtgttcagggtgttcagggtgttc 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2135 GTATCACCGRACAGCGGTATATGGCGTCAACCTGCGCGGTAGTACCCCTTCTTATCG 2076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

QY 729 gtatcggtaccgcagctatcgtggtggccttgattgacactctttaaataatcaagcagttg 788
Db 2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCAATATTGGCTACGGCGAAACAAAGTCTACT 2016
QY 789 cgttggcgctgcaggtcttcttgggtgtgttcttctattgatgcctcagatatggtcatgt 848
Db 2015 CCTTTGTTGGCAAGTATTTTACCTTCATGCAAAACCATCCGTCACGGGAATTGATT 1956
QY 849 tcttgggtgtagtggatgtttacctcttcttcacatcgcattcggcgagcaggttctatggcc 908
Db 1955 TCACCTCTGGGCCACCGTATTTGGCGGTGTCTATTCGCCATCGTTCGGCATTTTGTGGTA 1896
QY 909 tttacttggctgcgcgaacgcagcagcatgatccagatcacaacgcgtctccagtgccgtg 968
Db 1895 CGGTGATGCTTATTCATCACCGCTAAAGCTCAGCGAGCGAGGTGCCCC----- 1842
QY 969 caggaaacgcaaacgcaagcagaagcagcccgacagatttcaaacgattccaccatca 1028
Db 1843 -----GAAGAGAAAACACAGAGGTTATTACACCATCGAGCAGGCGGTA 1798
QY 1029 tccaggcaactttgacccgtgaagctattgcactgagcagcgtcagcagatgcatgtttg 1088
Db 1797 TCGTTCACCGATGACGGGAGAGATTGTGTGCTCATTCACGTCGTGATACACCGTTTG 1738
QY 1089 ccagcgaaagcttgcctgcggcggttgccatcgctcccaacaaaggcgagttagttctc 1148
Db 1737 CCACTGGCCTGTGGTAAGGTATTGCCATTCGCCCTCGGTGGTGTGAAGTCGCTCTC 1678
QY 1149 cgtgagtgaaagattggtggcattcccatctgcccattgcttccgagttcgcacca 1208
Db 1677 CGGTTCGGGTGCAATGCTTCGTTGCGCCACATTACA-----CGCATTTGGCATTG 1624
QY 1209 aggtcagagatggttccaatgtggtgatcttcttgatgcacattggtttcgcacagtaaac 1268
Db 1623 AGTCAGATGATG-----TGTGGAGATCCTGATTTCATGTCGTATCGACACCGTAAAC 1570
QY 1269 tcaacgcagcactttaaccgcctgaagaagcagcagcagatgaagtcaagcagggagc 1328
Db 1569 TGACGGCAAAATCTTTTCGCTCAGCTCAACGTGGGTGACAAAGTCAATACAGGCGATC 1510
QY 1329 tctgtgtgaattcgatattgatgccattaaaggctcaggttatagagtaaacacgcgca 1388
Db 1509 GGCTGATTTCTTTTGATATCCCTGCTATTTCGCGAGCGCGGATTGTATCTGACGACCGCG 1450
QY 1389 ttgtgttgaattacagaaacgcgacctgtaaacattcagttgttggcgaaattg 1448
Db 1449 TATTAAATCAATAAGCATGATTTTACGGACGATATTACCCACGCGATACCGGCGCAGATAA 1390
QY 1449 aagcggagcacaacctgctcaacgtc 1474
Db 1389 GCGCAGGTGAACCGCTGTATTCCATC 1364

```

RESULT 5

```

US-08-920-828-17/c
; Sequence 17, Application US/08920828
; Patent No. 5853996
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
; US-08-920-828-17

```

```

Query Match 7.2%; Score 109.6; DB 2; Length 3615;
Best Local Similarity 46.3%; Pred. No. 2.3e-23;
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;

```

```

QY 312 tcaactcagtgagcgttgcctcactcaccggattccttacattacatcgccatggcccg 371
Db 2495 TCACACATTTGCTATGCTGATGTTATACACCCGTCACCTTTCTGCTGGTGGGCGCG 2436
QY 372 caatgcgctgggtggcgagtgctggtgcacacggctctacaggagactttatgattcgggt 431
Db 2435 TATCAACCTGGATAAGCGAACTGATTGCCGCGGTATCTCTGGCTTTATCAGGCGGTT 2376
QY 432 gtccagtcggcgctcgtcttcgttgcgttgccttactactacccaatcgatcactggtctgc 491
Db 2375 CTGCAATTTGGGCGCGGTAAATGGCGGCTCTGGCAAACTCTTCGTCATCTTCGGACTGC 2316
QY 492 accagtccttcgcccaattgagctggtttaaccagggtggaatccttcatcttcg 551
Db 2315 ACTGGGCGCTGGTGGCGGTGTATCANATACTTACCGTCTGGGTACGACACCATGA 2256
QY 552 caacggcatctatg---gctaataatcgccaggggtgcggcatgtttggcagtggtcttcc 608
Db 2255 TCCCGCTGTTAATGCCCGCATTTATGCGCAGGTGCGGGCGCGCTCGGGGTCTTCTCTCT 2196
QY 609 tggcgaagagtgaaagctcaaggcccttcagagtgcttcaggtgtctccgctgttcttg 668
Db 2195 GCGAACCGGATGCGCAGAAAAAGTGTGCGGGATCAGCGCGGCTTACGAGTCTGTTG 2136
QY 669 gtattcagagcctgcgactctcgtgtgaaaccttcgcctgcgctgagcgttcttcacg 728
Db 2135 GTATCACCGAACCGCGGTATATGGCGTCAACCTCGCGCTAAGTACCCCTTTGTTATCG 2076
QY 729 gtatcggtaccgcagctatcggtggcgcttgatgcactctttaaatacaaggcagttg 788
Db 2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCAATATTGGCTACGGCGAAACAAAGTCTACT 2016
QY 789 cgttggcgctgcaggttcttcttgggtgtgttcttctatgatgtccagatatggtcatgt 848
Db 2015 CCTTTGTTGGCAAGTATTTTACCTTCATGCAAAACCATCCCGTCAACGGGAATTGATT 1956
QY 849 tcttgggtgtagtggatgtttaccttcttcacatcgcattcggcgagcaggttctatggcc 908

```

Db 1955 TCACGGCTGGCCAGCGTTATTGGCGGTGTCATTTGCCATCGGTTGGCATTGTTCGGTA 1896  
Qy 909 tttaactgtgttcgcgcgaacgagcagcattgtatccagatgcaacccgctgtctccagtgctg 968  
Db 1895 CGGTGATGCTTTCATTTCATCCAGCGGTAACCTACGCCAGCGAGGTTGCC----- 1842  
Qy 969 caggaacgacaaagcgaagcagaagcaccgcagaaattttcaaacgattccaccatca 1028  
Db 1843 -----GAAGAGAAACACACAGAGGTTATTACACCACTGAGCAGGCGGTA 1798  
Qy 1029 tccaggcacctttgaccggtgaagctattgacctgagcagcgtcagcagcagcagcagcagc 1088  
Db 1797 TCTGTTCACCGATGACGGAGAGATTGTGCTCATTCATTCAGTGCCTGATACACAGTTG 1738  
Qy 1089 ccagcggaagctgtgctggtcggtgtgcatgctcccaaccaagggcgagttgtctc 1148  
Db 1737 CCAGTGGCCTGTGGGTAAAGGTATTGCCAATCTGCCCTCGGTTGGTGAAGTGGCTTCTC 1678  
Qy 1149 cgggtgagtgaaagattggtggtgcatcccatctgcatctggttccagtcagtcgaccca 1208  
Db 1677 CGGTTGCCGGGTGCAATTGCTTCTGTTGTCGCACATTAACA-----CGCCATTGGCAATTG 1624  
Qy 1209 aggtcgagtggttcccaattggtatctgtgatgacattggtttcgacacagtaaac 1268  
Db 1623 AGTCAGATGATGG-----TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAANA 1570  
Qy 1269 tcaacggcagcacttaacccgctgaagaagcagggcgatgaagtcaaaacagcaggggagc 1328  
Db 1569 TGGACGGCAAAATCTTTTCCGCTCAGCTCAACGTTGGGTGACAAAGTCAATACAGCGGATC 1510  
Qy 1329 tgcgtgtgaattcgatattgatccattaaaggtgaggttatgaggttaacacacgcga 1388  
Db 1509 GGCTGATTTCTTTTATATATCCTCTGTAFTCGCAGCGCGGATTTGATCTGACGACGCGG 1450  
Qy 1389 ttgtgttgcgaattacaagaacacgggacacctgaacacactacggtttggcgcaaatg 1448  
Db 1449 TATTAACTAGTAAGCATGATTTTACGGAGCATTTACCCACCGACGCGCGGAGATAA 1390  
Qy 1449 aagcgggagcacaacctgctcaagtc 1474  
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

## RESULT 6

US-08-673-190A-3/c  
; Sequence 3, Application US/08673190A  
; Patent No. 5985668  
; GENERAL INFORMATION:  
; APPLICANT: Mattes, Ralf  
; APPLICANT: Klein, Kathrin  
; APPLICANT: Stegmaier, Sabine  
; TITLE OF INVENTION: Sucrose Metabolism Mutants.  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673.190A  
; FILING DATE: 27-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Forman, David S.  
; REGISTRATION NUMBER: 33,694  
; REFERENCE/DOCKET NUMBER: 06473.0001-00000

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)408-4000  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 465 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORGANISM: Protaminobacter rubrum  
; US-08-673-190A-3

Query Match 6.4%; Score 97.8; DB 2; Length 465;

Best Local Similarity 53.5%; Pred. No. 2.8e-20;

Matches 204; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

Qy 145 cgccaccatgctgcggcggaatgccaatgtgtccctgtttggttttagatgttgcaca 204  
Db 441 CGCCTGGGGCTCGCGGGCGGTTTCCACACCATGAATTTCTTCGACTCGATATCGCCAT 382  
Qy 205 agccggttacagggcaccgctctctgtctgctggtgttcttggattctggcaacgat 264  
Db 381 GATCGGTATCAGAGTACGGTGTTCGGGTGCTGCTACGGGTGGTGTATGAGCGTGT 322  
Qy 265 cgagaagtctctcacagcgactcaaggcgactgagggcgactgagactctctgactccagtgct 324  
Db 321 GGAANAACAGTCGCGAGAGTGATCCCAACGCGCTGGACCTGATCTCGACCATTCCT 262  
Qy 325 gacgttgcgtcacccgagattccttacattcatcgccattggcccgacgaatgcgctgggt 384  
Db 261 GACGCTGGTATCTCCGGCTTCGTCGCGCATGCTGTTTATCGCCCGCGCGCGCACACT 202  
Qy 385 gggcgatgctgcgcacacggtctacaggggactttatgatttcggtggtccagtcggcgg 444  
Db 201 GGGTGACGGCATCTCTTCGTCGCTCAGACACGCTGATTGCCACACCGCGTGTGGTTGCGCGG 142  
Qy 445 tctgctcttcggtctggtctactcaccgaatgctcatcactggtctctgacacagtccttccc 504  
Db 141 ATTCTGTTCGGCGGCTGTATTTCGGCCATGCTATCACCAGGCATTACACACAGCTTCCA 82  
Qy 505 gccaatgagctggagctgtt 525  
Db 81 GCGCGTAGAGGGGGGCTGCT 61

## RESULT 7

US-08-673-190A-6  
; Sequence 6, Application US/08673190A  
; Patent No. 5985668  
; GENERAL INFORMATION:  
; APPLICANT: Mattes, Ralf  
; APPLICANT: Klein, Kathrin  
; APPLICANT: Stegmaier, Sabine  
; TITLE OF INVENTION: Sucrose Metabolism Mutants.  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673.190A  
; FILING DATE: 27-JUN-1996

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 06473.0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Enterobacter species
; US-08-673-190A-6

Query Match          4.7%; Score 71.6; DB 2; Length 357;
Best Local Similarity 53.1%; Pred. No. 2.6e-12;
Matches 152; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 183 tggttggttagatgtgcccacagccggttacccaggcaccgtgctctcctgtgctgtg 242
Db 67 TCTTCGCATCAGTAGGCGCATGATCGGCTACACAGGCGACCGTCTCCCGGTGCTGCTGG 126
QY 243 ttcttggtatctggcacaagatcgagaagttcctgcacaaagcactcaaggcgactgcag 302
Db 127 CGGTGTGGTTATAGCATGGTGCAGAAAGCGGTGCGCSCGTTATCCCTGACGCGCTGG 186
QY 303 actctctgatcactcagtgctgacgttgctgctcaccggattccttaccattcatcgcca 362
Db 187 ACCTGATCTCTACCCCGCTCTGACGGTGATTATCTCCGGCTTTATCGCCCTCTGCTGA 246
QY 363 ttggccacgaatgcgtgggtggcgatgctgctgcacacacggtctcacaggggacttatg 422
Db 247 TCGGCGCGCGCGCTGCGCGCTCGGCGACGGTATTTCGTTATCCCTACGACCGCTTATCA 306
QY 423 attcgggtggtccacgctggcggtgctgctcttcggtctggtctactc 468
Db 307 GCCACGCGCGCTGCTGCGCGGCTGCTGTCGCGCGCGCTTACTC 352

RESULT 8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22131-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
```

```
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match          2.8%; Score 43.2; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.0078;
Matches 30; Conservative 207; Mismatches 185; Indels 0; Gaps 0;

QY 147 ccaccatggtcggtcggtgcaaatgccaatgtggtccctgtgtttagattgcccaga 206
Db 1034 CGAGCTTGGCTCGAGGTCGAGGAGCTGCGATVYYYYYYYYYYYYYYYYYYYYY 1093
QY 207 ccggttacccaggcaccgtgcttccctgctgctggtgttcttctgattctgcaacatcg 266
Db 1094 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1153
QY 267 agaagttcctgcacagcgactcaaggcgactgcagactcctcctgactcactccagtctga 326
Db 1154 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1213
QY 327 cgtgtgtctcaccggtatccttaccattcctcattcgcattgcccagaatcgctgggtg 386
Db 1214 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1273
QY 387 gcgatgtgtggcacacggtctacaggagacttatgattcgtggtccagtcggcggtc 446
Db 1274 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1333
QY 447 tgctctcgtgctggtctactcaccaatgctcactggtgctgacacagtccttccgc 506
Db 1334 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1393
QY 507 caattgagctggagctgtttaaccagggtggatcccttcattcttcgaacggcatctatg 566
Db 1394 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1453
QY 567 ct 568
Db 1454 CT 1455

RESULT 9
US-08-936-165A-153
; Sequence 153, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
```



```

RESULT 12
US-08-460-751-1/c
; Sequence 1, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:

```

RESULT 13  
US-08-658-136-4/c  
; Sequence 4, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,136  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400

```

; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 135..13040
US-08-658-136-4

Query Match 2.1%; Score 32.8; DB 3; Length 14060;
Best Local Similarity 48.9%; Pred. No. 18;
Matches 88; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 235 gctggtgtttcttgatttgcaacgactgagaagtctctgcacacgactcaagg 294
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7193 GTTGGTGGCTCTCTCCCTTGGCGCGCGCTTCCACACGGTCAGGCTGAAGGTGTACTCCAC 7134

QY 295 cactgcgaactctctgatactccagtcgactgctgacgttctgctcaacggattccttacatt 354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7133 GCCAGCGCCGACGCGCTCCCGTGGAAATGGTACCGTCTGCTCCCGGGGCCCAAGTT 7074

QY 355 catcgccattggccagcaaatgcgtgggtggcgatgtgctgacacggtctacagg 414
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7073 CAGCGCACACCGCGCACCCCTCCCTCTGTGTGGAAGCCACACAGGCCCTAGTGGAAACTGAG 7014

RESULT 14
US-09-135-994-1
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-1

Query Match 2.1%; Score 32.6; DB 4; Length 477;
Best Local Similarity 60.9%; Pred. No. 2.7;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 920 cgccgaacggcagcattgatccagatgcaacgctctccagtcctcaggagcagcacc 979
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 108 cggagcagcgccgcgcggcagcagcagcagcagcagcagcagcagcagcagcagca 167

QY 980 aaagcgaagcagaagcaccgcgcagaa 1006
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 168 gcagcagcagcagcagcagcagcagca 194

RESULT 15
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:16:59 ; Search time 234.58 Seconds  
(without alignments)  
11176.265 Million cell updates/sec

Title: US-09-604-231-1

Perfect score: 1527

Sequence: 1 ctcatgcatctgcgcgtt.....gttgaacaccttgagtggtcg 1527

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID       | Description         |
|------------|--------|-------------|---------|-------------|---------------------|
| 1          | 1527   | 100.0       | 1527    | 22 AAF31528 | C-glutamicum phosph |
| 2          | 1515   | 99.2        | 349980  | 22 AAH68533 | C-glutamicum codin  |
| 3          | 1492   | 97.7        | 1983    | 22 AAH67869 | C-glutamicum codin  |
| 4          | 1412.6 | 92.5        | 5969    | 22 AAF32543 | Brevibacterium lac  |
| 5          | 1107.4 | 72.5        | 1109    | 22 AAF31529 | C-glutamicum phosph |
| 6          | 190.6  | 12.5        | 2913    | 22 AAH54100 | S-epidermidis gen   |
| 7          | 151.2  | 9.9         | 5840    | 20 AAX12968 | Enterococcus faeca  |
| 8          | 132.2  | 8.7         | 2944528 | 24 ABA03041 | Listeria monocytog  |
| 9          | 126.6  | 8.3         | 3895    | 19 AAV52334 | Streptococcus pneu  |

|    |       |     |         |             |                     |
|----|-------|-----|---------|-------------|---------------------|
| 10 | 113.4 | 7.4 | 1884    | 23 AAS55732 | Streptococcus pneu  |
| 11 | 108   | 7.1 | 3515    | 15 AAQ55752 | Escherichia coli g  |
| 12 | 107.8 | 7.1 | 9769    | 19 AAV52163 | Streptococcus pneu  |
| 13 | 97.8  | 6.4 | 465     | 18 AAT47502 | Partial P. rubrum   |
| 14 | 86.2  | 5.6 | 30246   | 18 AAT47367 | Staphylococcus aur  |
| 15 | 81.6  | 5.3 | 2944528 | 24 ABA03041 | Listeria monocytog  |
| 16 | 76.4  | 5.0 | 465     | 18 AAV75628 | Staphylococcus aur  |
| 17 | 74.2  | 4.9 | 474     | 20 AAX13576 | Enterococcus faeca  |
| 18 | 71.6  | 4.7 | 357     | 18 AAT47505 | Partial Enterobact  |
| 19 | 70.4  | 4.6 | 9797    | 20 AAX13487 | Enterococcus faeca  |
| 20 | 67    | 4.4 | 8494    | 19 AAV52296 | Streptococcus pneu  |
| 21 | 66    | 4.3 | 1947    | 22 AAH81338 | Escherichia coli p  |
| 22 | 66    | 4.3 | 1947    | 23 AAS52355 | Escherichia coli p  |
| 23 | 65.2  | 4.3 | 350     | 23 AAS87225 | E. coli DNA for ce  |
| 24 | 61.6  | 4.0 | 1319    | 23 AAS52023 | DNA encoding novel  |
| 25 | 61.6  | 4.0 | 1455    | 23 AAS54537 | Staphylococcus aur  |
| 26 | 61.2  | 4.0 | 22934   | 23 AAS59613 | Staphylococcus aur  |
| 27 | 60.4  | 4.0 | 2550    | 22 AAH54979 | Propionibacterium   |
| 28 | 60    | 3.9 | 6285    | 20 AAX13352 | S-epidermidis gen   |
| 29 | 59.2  | 3.9 | 2049    | 22 AAH68426 | Enterococcus faeca  |
| 30 | 59.2  | 3.9 | 2172    | 22 AAF31542 | C-glutamicum codin  |
| 31 | 59.2  | 3.9 | 349980  | 22 AAH68528 | C-glutamicum phosph |
| 32 | 58.4  | 3.8 | 1368    | 23 AAS56059 | Salmoneilla typhi D |
| 33 | 56.8  | 3.7 | 29555   | 18 AAV74517 | Staphylococcus aur  |
| 34 | 54.2  | 3.5 | 428     | 22 AAF31544 | C-glutamicum phosph |
| 35 | 53.2  | 3.5 | 2996    | 22 AAH54445 | S-epidermidis gen   |
| 36 | 53.2  | 3.5 | 3081    | 22 AAH54946 | S-epidermidis gen   |
| 37 | 53.2  | 3.5 | 3932    | 22 AAH54056 | S-epidermidis gen   |
| 38 | 51.8  | 3.4 | 7156    | 20 AAX12966 | Enterococcus faeca  |
| 39 | 51.4  | 3.4 | 1906    | 20 AAX13595 | Enterococcus faeca  |
| 40 | 51.4  | 3.4 | 2882    | 19 AAV52273 | Streptococcus pneu  |
| 41 | 51    | 3.3 | 2037    | 23 AAS52021 | Staphylococcus aur  |
| 42 | 51    | 3.3 | 2046    | 23 AAS54535 | Staphylococcus aur  |
| 43 | 51    | 3.3 | 29555   | 18 AAV74517 | Staphylococcus aur  |
| 44 | 49.2  | 3.2 | 796     | 18 AAV74736 | Staphylococcus aur  |
| 45 | 46.8  | 3.1 | 2529    | 23 AAS54215 | Pseudomonas aerugi  |

#### ALIGNMENTS

RESULT 1  
ID AAF31528 standard; DNA; 1527 BP.

XX AAF31528;

DT 09-APR-2001 (first entry)

XX C-glutamicum phosphoenolpyruvate DNA #1.

DE Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.

XX Corynebacterium glutamicum.

XX WO200102583-A2.

XX 11-JAN-2001.

XX 27-JUN-2000; 2000WO-IB00973.

XX 01-JUL-1999; 99US-0142691.

PR 23-AUG-1999; 99US-0150310.

PR 03-SEP-1999; 99DE-1042095.

PR 03-SEP-1999; 99DE-1042097.

XX (BADI ) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-080989/09.

XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:  
PT

PT sugar phosphotransferase system proteins or their portions, useful for  
PT typing or identifying C. glutamicum or related bacteria, and as markers  
XX for transformation -  
XX  
PS Claim 3; Page 98-101; 144pp; English.  
XX

CC The present invention relates to Corynebacterium glutamicum  
CC phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.  
CC The PTS nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria, the typing or identification of C. glutamicum or  
CC related bacteria, as reference points for mapping C. glutamicum genome,  
CC and as markers for transformation.  
XX  
SQ Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;

|                       |     |  |               |           |              |
|-----------------------|-----|--|---------------|-----------|--------------|
| Query Match           |     | 100.0%;  | Score 1527;   | DB 22;    | Length 1527; |
| Best Local Similarity |     | 100.0%;  | Pred. NO. 0;  |           |              |
| Matches 1527;         |     | Conservative 0;  | Mismatches 0; | Indels 0; | Gaps 0;      |
| Qy                    | 1   | ctcatggcatctgcgcgttcgcttcttcgagtggttgcagtggttggtttccacgcgaaccaag | 60            |           |              |
| Db                    | 1   | ctcatggcatctgcgcgttcgcttcttcgagtggttgcagtggttggtttccacgcgaaccaag | 60            |           |              |
| Qy                    | 61  | cggttcggcgcaatagttcttcggcgccgctattggtatgagcgatggtgttcccgag       | 120           |           |              |
| Db                    | 61  | cggttcggcgcaatagttcttcggcgccgctattggtatgagcgatggtgttcccgag       | 120           |           |              |
| Qy                    | 121 | cttggtgaacggtcacagctggtgcgcacacataggttcggtggcgcaaatgccaatggtgctc | 180           |           |              |
| Db                    | 121 | cttggtgaacggtcacagctggtgcgcacacataggttcggtggcgcaaatgccaatggtgctc | 180           |           |              |
| Qy                    | 181 | cctgttggttagatgttgccaaagcgggttacaggcgacccgtgtctctgtcgtggt        | 240           |           |              |
| Db                    | 181 | cctgttggttagatgttgccaaagcgggttacaggcgacccgtgtctctgtcgtggt        | 240           |           |              |
| Qy                    | 241 | gggttcttggtgattggtgcaacgacgagaaagtctctgcacagcgactcaaggcgactgc    | 300           |           |              |
| Db                    | 241 | gggttcttggtgattggtgcaacgacgagaaagtctctgcacagcgactcaaggcgactgc    | 300           |           |              |
| Qy                    | 301 | agaattctgatactccagtgctgaagtgctgctcacaggatctcttaacatcatcgc        | 360           |           |              |
| Db                    | 301 | agaattctgatactccagtgctgaagtgctgctcacaggatctcttaacatcatcgc        | 360           |           |              |
| Qy                    | 361 | cattggccagcaatcgctggtggcgatgtgctgacacaggtctacagggaactta          | 420           |           |              |
| Db                    | 361 | cattggccagcaatcgctggtggcgatgtgctgacacaggtctacagggaactta          | 420           |           |              |
| Qy                    | 421 | tgatttcggtggtccagtcggtgctgctcttcggtgctgctactacccaatcgctcat       | 480           |           |              |
| Db                    | 421 | tgatttcggtggtccagtcggtgctgctcttcggtgctgctactacccaatcgctcat       | 480           |           |              |
| Qy                    | 481 | cactgggtgacacagtccttcgcccaattgagctgagctgtttaaaccagggtggtatc      | 540           |           |              |
| Db                    | 481 | cactgggtgacacagtccttcgcccaattgagctgagctgtttaaaccagggtggtatc      | 540           |           |              |
| Qy                    | 541 | cttcaatcttcgcaacggcatctatggttaataatcgcgccagggtgcgcgcatgttggcagt  | 600           |           |              |
| Db                    | 541 | cttcaatcttcgcaacggcatctatggttaataatcgcgccagggtgcgcgcatgttggcagt  | 600           |           |              |
| Qy                    | 601 | gttctctcgccagagtagtgaagagctcaaggccttcgaggtgttcaggtgctctccgc      | 660           |           |              |
| Db                    | 601 | gttctctcgccagagtagtgaagagctcaaggccttcgaggtgttcaggtgctctccgc      | 660           |           |              |
| Qy                    | 661 | tgttcttggtattacggagctcgatcttcggtgtgtaaccttcgctgcgtggtggcggtt     | 720           |           |              |
| Db                    | 661 | tgttcttggtattacggagctcgatcttcggtgtgtaaccttcgctgcgtggtggcggtt     | 720           |           |              |
| Qy                    | 721 | cttcaatcggtatcggtaccgcagctatcggtggcggtttgattgacatctttaaatacaa    | 780           |           |              |
| Db                    | 721 | cttcaatcggtatcggtaccgcagctatcggtggcggtttgattgacatctttaaatacaa    | 780           |           |              |

|    |      |   |      |  |  |
|----|------|---|------|--|--|
| Qy | 781  | ggcagttgcgttggtggcgctgcaggtttcttgggtgtgtgtttctattgatgtccagatat  | 840  |  |  |
| Db | 781  | ggcagttgcgttggtggcgctgcaggtttcttgggtgtgtgtttctattgatgtccagatat  | 840  |  |  |
| Qy | 841  | ggtcatgttcttggtggtgcagttgttaccttcttcattcgcattcgcgcagcgattgc     | 900  |  |  |
| Db | 841  | ggtcatgttcttggtggtgcagttgttaccttcttcattcgcattcgcgcagcgattgc     | 900  |  |  |
| Qy | 901  | ttatggcctttacttggttgcgcacacggcagcatgtatccagatgcaacgcgtgctcc     | 960  |  |  |
| Db | 901  | ttatggcctttacttggttgcgcacacggcagcatgtatccagatgcaacgcgtgctcc     | 960  |  |  |
| Qy | 961  | agtgcctgcaggaacgacccaagcgaagcagaagcaccgcagaaattttcaaacgattc     | 1020 |  |  |
| Db | 961  | agtgcctgcaggaacgacccaagcgaagcagaagcaccgcagaaattttcaaacgattc     | 1020 |  |  |
| Qy | 1021 | caccatcatcaggcacacctttgacccgtggaagctatgtcactgagcagcgtcagcgatgc  | 1080 |  |  |
| Db | 1021 | caccatcatcaggcacacctttgacccgtggaagctatgtcactgagcagcgtcagcgatgc  | 1080 |  |  |
| Qy | 1081 | catgtttgcacgaggaagcgttgctcggtggttcgcatctcccaacccaaggcgagtt      | 1140 |  |  |
| Db | 1081 | catgtttgcacgaggaagcgttgctcggtggttcgcatctcccaacccaaggcgagtt      | 1140 |  |  |
| Qy | 1141 | agtttcctcggtgagtggaagattgtggtggcattcccatctggccatgcttttcgcagt    | 1200 |  |  |
| Db | 1141 | agtttcctcggtgagtggaagattgtggtggcattcccatctggccatgcttttcgcagt    | 1200 |  |  |
| Qy | 1201 | tcgcacaaagcgtgagatggttccaatgtggtatcttgatgacattggtttcgacac       | 1260 |  |  |
| Db | 1201 | tcgcacaaagcgtgagatggttccaatgtggtatcttgatgacattggtttcgacac       | 1260 |  |  |
| Qy | 1261 | agtaaacctcaacgcgcgcacatttaaccgctgaagcagcagcgatgaagcgaagc        | 1320 |  |  |
| Db | 1261 | agtaaacctcaacgcgcgcacatttaaccgctgaagcagcagcgatgaagcgaagc        | 1320 |  |  |
| Qy | 1321 | aggggagctgctgtgtaattcgatgtgcatcattgaagcctgaagcgtcaggttatgaggaac | 1380 |  |  |
| Db | 1321 | aggggagctgctgtgtaattcgatgtgcatcattgaagcctgaagcgtcaggttatgaggaac | 1380 |  |  |
| Qy | 1381 | cacgcgattgttcttcgaattacaagaaacccgacccgtgaacacattacggtttggg      | 1440 |  |  |
| Db | 1381 | cacgcgattgttcttcgaattacaagaaacccgacccgtgaacacattacggtttggg      | 1440 |  |  |
| Qy | 1441 | cgaattgaagcggagccacacctgcacacgtcgaagcgaagcgaagcgaagcgaagc       | 1500 |  |  |
| Db | 1441 | cgaattgaagcggagccacacctgcacacgtcgaagcgaagcgaagcgaagcgaagc       | 1500 |  |  |
| Qy | 1501 | accataagttgaaaccttgagttctg                                      | 1527 |  |  |
| Db | 1501 | accataagttgaaaccttgagttctg                                      | 1527 |  |  |

RESULT 2  
AAH68533/c  
ID AAH68533 standard; DNA; 349980 bp.  
XX  
AC AAH68533;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 7068.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX  
XX EPI108790-A2.  
XX  
PD 20-JUN-2001.  
XX



PF 18-DEC-2000; 2000EP-0127688.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI; 2001-376931/40.  
XX  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX Disclosure; SEQ ID NO: 7068; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

Query Match 99.2%; Score 1515; DB 22; Length 349980;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1526; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 cttatggcatcgcgcgttcgcttctgcccagtggttggttcacccgcaacccag 60  
Db 111326 CTCATGGCATCTGCGCGTTCGCTCTCTGCGAGTGTGGTGGTTCACCGCAACCAAG 111267

QY 61 cgtttcggcggcaatgagttcctctggcgcgcgcgtatgtatggcgatgggttcccgag 120  
Db 111266 CGTTTCGGCGGCAATGAGTTCCTGCGCGCGC-GTATTGCTATGGCGATGGTTCGCCGAG 111208

QY 121 cttggtgaacggtacgacgtggccgcgcacccatggctgcggcgcaaatgcaatgtggttc 180  
Db 111207 CTTGGTGAACGGGTACGAGTGGCGCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTC 111148

QY 181 cctgtttggttagatgttgcccaagccggttaccaggcaccgctgcttcctggtggtgt 240  
Db 111147 CCTGTTTGGTTAGATGTTGCCAAGCCGCTTACCAGGCAACCGTCTTCTGTGCTGGT 111088

QY 241 ggtttcttgatctcggcaacgatcgagagtgcttcctgcacaaagcgactcaaggcgactgc 300  
Db 111087 GGTTCCTTGGATTTCTGGCAACGATCGAGAAGTTCCTGCAACAGCGACTCAAGGGCACTGC 111028

QY 301 agacttcctgatcaactccagtgctgaactgtgctgtcaccggatctcctacattcatcgc 360  
Db 111027 AGACTTCCTTGATCACCTCCAGTGTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 110968

QY 361 catgtggccacgaatcgctgggtggcgatgtgctggcacacgctctacagggaacttta 420  
Db 110967 CATTGGCCCGCAGAAACGCGTGGGTGGCGCATGTGCTGGCACACGCGTCTACAGGGACTTTA 110908

QY 421 tgatttcggttccagtcggcggtgtctctctcgtctggtctactcaccatcgatcat 480  
Db 110907 TGATTTTCGGTGTGCCAGTCCGGCGGTCTGTCTCTTCGGTCTGCTACTACTACCAATCGTCAT 110848

QY 481 cactggtctgcaccagtcctcccgcccaattgaagctggagctgttttaaccagggtgatc 540  
Db 110847 CACTGTGCTGACACGACTCCTTCCCGCCAATTGAGCTGGAGAGTGTTTAAACAGAGGGTGAATC 110788

QY 541 cttcatcttcgcaacggcatctatggtctaataatcgccagggtgctggcgatgtttggcagt 600  
Db 110787 CTTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCAGGGTGGCGCATGTTTGGCAGT 110728

QY 601 gttcttcctcggcgaagagtgaagagctcaaggccttgagggtgcttcaggtgtctccgc 660  
Db 110727 GTTCTTCCTTGGCAGAGAGTGAAAGGCTCAAGGGCTTGCAGGTGCTTCAGGTGCTCTCCGC 110668

QY 661 tttctctgtattacagagacctgcgactctcgtgtgaaccttcgctcgcctcgctggccggt 720  
Db 110667 TGTCTTGTGATTTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCTGCGCTGGCGGCTT 110608

QY 721 cttcatcggtatcgttaccgcgactatcggtggcgctttgattgactctcttaatatcaa 780  
Db 110607 CTTTCATCGGTATCGGTACCGCAGCTATCGTGGCGCTTTGATTCGCACTCTTTAATATCAA 110548

QY 781 ggcagttgcgttggcgctgcagggttctctgggtgtgtgtttctattgatgctccagatat 840  
Db 110547 GGCAGTTCGCTTGGCGGCTGCGAGTTCCTTGGGTGTGTTTCTTATTTCATGCTCCAGATAT 110488

QY 841 ggtcatgttcttggtgtgtgcagttgttaccttcttcacgacttcggtgcgacgattgc 900  
Db 110487 GGTTCATGTTCTTGGTGTGTCAGTGTGTACCTTCCTTCATCGCATTCGGCGCAGCATTCG 110428

QY 901 ttatggccttacttggtttcgcccgaacggcagcattgatccagatgcacacgctcctcc 960  
Db 110427 TTATGGCCCTTACTTGGTTCGGCGCAACGGCAGCATTTGATCCAGATGCAACCGCTCTCC 110368

QY 961 agtgctcgcaggaacacccaaagccgaagcagcagcagcagcagcagcagcagcagcagc 1020  
Db 110367 AGTGCTCTGAGGAACGACCAAGCCGAGCAGAACCCGCGCAGAAATTTTCAAACGATTC 110308

QY 1021 caccatcatccaggcaccctttgacggtgaagctattgactgagcagcagcagcagcagc 1080  
Db 110307 CACCATCATCCAGGCACTTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTACGCGATGC 110248

QY 1081 catgtttgcagcggaagcgttggctcggcgcttgccatgctcccaaccaagggcgagtt 1140  
Db 110247 CATGTTTGCAGCGGAAGCTTGGCTCGGCGTTCGCTCCCAACCAAGGGCGAGTT 110188

QY 1141 agttctccggtgagtggaagatgtgtggtggtcattcccatctggccatgcttccgagtt 1200  
Db 110187 AGTTTCTCGGTTGAGTGGAAAGATTGTGTGGCATTCCTCATCTGGCATGCTTTTCGCAGT 110128

QY 1201 tcgcaccaaggtgagatggttccaatgtggtatcttgatgcacattggttctgcacac 1260  
Db 110127 TCGCACCAAGGCTGAGGATGGTTCCAAATGTGGATATCTTTGATGCATTTGGTTTCGACAC 110068

QY 1261 agtaaacctcaacggcacgcacactttaaccgctgaagaagcagggcgatgaagtcaaaagc 1320  
Db 110067 AGTAACCTCAAGGGCACGCACTTTAACCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC 110008

QY 1321 agggagctcgtgtgtgaattcgatattgatgccattaaggctgcaggttatgaggttaac 1380  
Db 110007 AGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTATGAGGTAAC 109948

QY 1381 cagcgcatgtgttttcgaattacaagaaacccggacccgtgaaacacattcaggtttggg 1440  
Db 109947 CAGCGCATGTGTGTTTCGAATTTACAAGAAACCGGACCTGTAAACACTTACCGTTTGGG 109888

QY 1441 cgaattgaagcgggagcacaacctgctcaacgtcgcaagaagaagcgggtgcccagcaac 1500  
Db 109887 CGAAATGAAGCGGGAGCCAACTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAAC 109828

QY 1501 accataagttgaacaccttgagtgttcg 1527  
Db 109827 ACCATAAGTTGAACCTTGAGTGTTCG 109801







QY 1199 gttcgaccaggctgagatggttccaatgtgatattgtgatgtgacacattggtttcgac 1258  
 Db 781 gttcgaccaggctgagatggttccaatgtgatattgtgatgtgacacattggtttcgac 840  
 QY 1259 acagtaaaactcaacgacgacactttaaccgctgaagaaagcagggcagatgaagcaaa 1318  
 Db 841 acagtaaaactcaacgacgacactttaaccgctgaagaaagcagggcagatgaagcaaa 900  
 QY 1319 gcaggaggagctggtgtaattcgatattgatgccattaaaggctcaggttataggta 1378  
 Db 901 gcaggaggagctggtgtaattcgatattgatgccattaaaggctcaggttataggta 960  
 QY 1379 accacgcccattgttttcgaattacaagaaacccgacctgaacacacttacggtttg 1438  
 Db 961 accacgcccattgttttcgaattacaagaaacccgacctgaacacacttacggtttg 1020  
 QY 1439 ggcgaattgaagcggggagccacacctgctcaacgctgcgaagaaagcgggtgcagca 1498  
 Db 1021 ggcgaattgaagcggggagccacacctgctcaacgctgcgaagaaagcgggtgcagca 1080  
 QY 1499 acacataaagtgaacacctgagttctg 1527  
 Db 1081 acacataaagtgaacacctgagttctg 1109

RESULT 6

AAH54100/C  
 ID AAH54100 standard; DNA; 2913 BP.  
 XX AC AAH54100;  
 XX DT 03-SEP-2001 (first entry)  
 XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.  
 XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 XX KW vaccination; endocarditis; ds.  
 XX OS Staphylococcus epidermidis.  
 XX PN WO200134809-A2.  
 XX PD 17-MAY-2001.  
 XX PF 09-NOV-2000; 2000WO-US30782.  
 XX PR 09-NOV-1999; 99US-0164258.  
 XX PA (GLAX ) GLAXO GROUP LTD.  
 XX PI Kimmerly WJ;  
 XX DR WPI; 2001-316495/33.  
 XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX PS useful for vaccinating against infections, e.g. endocarditis -  
 Claim 8; Page 1030-1031; 2188pp; English.  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX SQ Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;

Query Match 12.5%; Score 190.6; DB 22; Length 2913;  
 Best Local Similarity 55.0%; Pred. No. 6.2e-46;  
 Matches 468; Conservative 0; Mismatches 364; Indels 19; Gaps 4;  
 QY 5 tggcatctgcgcgcttccggtttcttccagtggttggtttccacgcaaccaagcgtt 64  
 Db 2690 TTGCAAAATGCACCTTTTACATATTATCCAAATACTTATTTGGTTTGTAGTCAGCTAAGCGAT 2631  
 QY 65 tcggcgcaatgagttccctcggcgccgctgattggttatggcgatggtgttcccgagcttg 124  
 Db 2630 TTGGTGGTAATCCTTATTTAGGTGCAGC-TCTAGGTATGATACTTGTTCATCTGGATTG 2572  
 QY 125 gtaacggctacgagctggcgcgccaccatgctgcggcgcaa---atgccaatgtgttcc 181  
 Db 2571 ATGAGTGCATATGATTTTCCAAAAGCTTTAGAGAAGAGAAAAGCTATTCCACACGGGAT 2512  
 QY 182 ctggttggttagatgttgcacacgggttaccagggcaccgtgcttccgtgctggtg 241  
 Db 2511 GTCTTTGGGTACATATTAAATGAAGTAGGTTATCAGGAGCAAGTATTACCTATGCTCGTA 2452  
 QY 242 gttcttggtattcggcaacgacgacgagatctcctgcacaaagcagcactcaaggcactgca 301  
 Db 2451 GCAACATATATTATTAGCTACGATTGAAAAATGTTTACGTAAAGCTATTTCACACTGTGTTA 2392  
 QY 302 gacttctgacacacagctgctgacgtgctgctcaccgagcttccattccatcagcc 361  
 Db 2391 GATAAATTTATGAGCCATTTATTAATCAATTTTATTTATACACATTTTAAACATTTTATTT 2332  
 QY 362 attggcccagcaatgcgctgggtggcgatgctgctggcacacgctctacagggactttat 421  
 Db 2331 GTAGGGCTGTCACTGTCAATTAGGTTATAGGTTATTTGTTATCTGATGATTGATTGTTATAT 2272  
 QY 422 gattcgggtggtccagtcggcggtgctgcttctcgggtcgtctactcacaacatcgctcatc 481  
 Db 2271 GAATTTGGTGGAGCTATTGTGGACTTATATTTGGTTTATTTATATGCGCAATCGTCATT 2212  
 QY 482 actggtctgcaccagtccttcccgccaattgagctg-----gagctgtttaac 529  
 Db 2211 ACAGGAATGCACCATAGCTTTATTGTCAGTTGAACACACATTAATTTGCTGATGCGACTAAA 2152  
 QY 530 cagggtggatccttcacatcttcgcaacgcatctatggctaatacgcacaggtgctggca 589  
 Db 2151 ACAGGTGGTTCAATTTATCTCCCAATCGCAACGATGTCAAATATTGCACAAGGTGGTGA 2092  
 QY 590 tgtttggcagtggtctt---tcctggcggaagagtgaaagcctcaaggccttgaggtgct 646  
 Db 2091 GCTTTAGCTGCATCTTCTTATCATTAAGCAAAATAAAAATTTAAAGGTGTGTGTCGCG 2032  
 QY 647 tcaggtgctccgcttcttctgattacggagccttcgcatcttcggttggaaccttcgc 706  
 Db 2031 CGGGGTATTTACGCTTTTACTAGGAATTTACAGAACACAGCAATGTTGGTGTCAATCTAAA 1972  
 QY 707 ctgctgctggcgttcttcacatcggtatcggtacccgagctatcggtgctgctgattgca 766  
 Db 1971 TTGAGATATCCATTTATAGTGTCTGTTGAGGATCAGGTATAGGTGCGGCTTATATTCA 1912  
 QY 767 ctctttaataacagcagcttggtggtgctgaggttctctctggtgtgtgtttctatt 826  
 Db 1911 TTCTTCAAAAGTAAACGATAGCGCTTGTGTACAGCTGGATTACCTGGATTATATCTATA 1852  
 QY 827 gatgctccaga 837  
 Db 1851 AATCCTACACA 1841

```

RESULT 7
AA12968
ID AAX12968 standard; DNA; 5840 BP.
XX
AC AAX12968;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:31.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN WO980555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
CC New isolated Enterococcus faecalis polynucleotides and polypeptides
CC - used to develop products for the detection of Enterococcus and for
CC use in vaccines for prevention or attenuation of Enterococcus
CC infection.
XX
PS Claim 1; Page 389-392; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 5840 BP; 1848 A; 944 C; 1256 G; 1779 T; 13 other;

Query Match 9.9%; Score 151.2; DB 20; Length 5840;
Best Local Similarity 48.5%; Pred. No. 4.9e-34;
Matches 525; Conservative 0; Mismatches 533; Indels 24; Gaps 3;

QY 329 ttgctgtccaccggtatcttaccattatccattgcccattggccagcaatgcgtgggtgggc 388
DB 1850 ttccattatggcacctgttacccttcttagctattgctctatcggtcgtgacgggtcatcggc 1909
QY 389 gatgtctggcacaggtctacagggaactttatgatttcgtgtccagtcggcggtctg 448
DB 1910 gatgtgtggacaaggatacaacgcaattacgctttttatgccaattatcgagggta 1969
QY 449 ctctctgggtctgtctactaccatcgtctatcactgctgtgtgacacagtccttcccaca 508
DB 1970 ttaatgggttcgtgtggcagctgtgtgtaattgttgatgcttgggtgggtctgtacca 2029

```

```

QY 509 attgagctggagctgtttaaccagggtggatcccttcattctcgcaacggcatctatggct 568
DB 2030 attatgatgttaacttaacacaagggtgcgatacagatggtaccgatgttattaccagcc 2089
QY 569 aatatccccagggtgcggcatgtttggcagctgttcttcctcggcgaagagtgaagaagctc 628
DB 2090 gttattgcacaagcggtgctgttagctgtcttttcttaacaaaaaaagtgaacta 2149
QY 629 aagggtcttcaggtcttcaggtctcccgctgttcttctgtattatcagagccctgcgctc 688
DB 2150 aaaggtttggcttcttcgaagtattacgactatttttggattactgaaccaactgta 2209
QY 689 ttccgtgtgaacctgcctgcgtgcgtgttcttcacggtatcgggtacgcgcagctatc 748
DB 2210 tatggcgtgactttaccattgaataaaaccatttattgcagctgttattgggtggcggtatc 2269
QY 749 ggtggcgtttgattgcactctttaataatcaaggcagtgctgtgggtggcgctgcaggtttc 808
DB 2270 ggtgtgcatttggctatgaatcaacgtgaaaaactttacgttttggctgttagtatg 2329
QY 809 ttgggtgtgtttctattgatgctccagatatggtcatgttcttctgtgtgtgcagttgtt 868
DB 2330 ttgagctgtccgtgttattcctcgcagagacaaaaagatactgcacogtatgattactgg 2389
QY 869 accttcttcacgcattcgcgcagcgtattgtcttattggcctttacttctgttgcgcgcaac 928
DB 2390 gcaattggtgccgaattgcctttatcatcgtgttctttaaactgttcttactgtt 2447
QY 929 ggcagcattgatccagatgcaacgcgtgctccagtcgctgcagtgcaaggaacccaaagcga 988
DB 2448 -----ttgaagatcaacctaatccagaacacgacactgaaaaaagacagaactgataag 2500
QY 989 gcagaacacccgcagaaattttcaaacgattccaccatcatccagga---cctttgacc 1045
DB 2501 atggtgcccactgtaaaaaacgaaatcaagaagacacaaaattttagcaagtcacttcaa 2560
QY 1046 ggtgaagctattgcactgagcagcgtcagcgtatgccatgttgcagcggaagcttggc 1105
DB 2561 ggtgaattttaccgctagaaaaagatacaagacctgtttgttgcaggtgtttagga 2620
QY 1106 tcgggtgttcacgtctcccaacaaagggcagttagtttctccggtgagtggaaagatt 1165
DB 2621 aaaggtgtgcaattgagcgcagactgaagcgaactgtatgacccgcagatggtgaaatc 2680
QY 1166 gtggtgcatctccatctgcccactgttcgacttgcagtcacaaaggtcagagatggttcc 1225
DB 2681 accacattatttccgcagggacatgctgtgtgcttgacgacacagagg----- 2729
QY 1226 aatgtgatattctgatgcacattggtttcgacacagtaaacctcaacggcacgcacttt 1285
DB 2730 -gcgtgaattattaatgcatattggcatgatacagtcgtaattagatggttaaaggcttt 2788
QY 1286 aaccgcctgaagaagcagggcgatgaagtcacaaagcaggggagctgctgtgtgaattogat 1345
DB 2789 gaattatcagtgaaacaagggtgattctgttaaaaaaggagattgttactttttag 2848
QY 1346 attgatccattaaagctgcaggttatgagtaaacccacgcogattgttcttgcgaattac 1405
DB 2849 attgctccattaaagaagctgttatccggttagttacacccgattgttgtaacacacg 2908
QY 1406 aa 1407
DB 2909 aa 2910

```

```

RESULT 8
ABA03041
ID ABA03041 standard; DNA; 2944528 BP.
XX
AC ABA03041;
XX
DT 05-FEB-2002 (first entry)

```

[illegible]









```
Qy 1346 attatgccattaaagctgcaggttatgagtaaacacgcgcgattgtgttttccaat 1402
Db 1747 toaacaaaaatcgtgcagctgagcttgatgatacaacaatggttatcacaat 1803

RESULT 11
AAQ55752/c
ID AAQ55752 standard; DNA; 3615 BP.
XX
AC AAQ55752;
XX
DT 25-SEP-1995 (first entry)
XX
DE Escherichia coli genomic probe EC-24.
XX
KW Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW K.pneumoniae; E.cloacae; clinical sample; ds.
XX
OS Escherichia coli.
XX
FN WO9401583-A.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-JP00936.
XX
PR 07-JUL-1992; 92JP-0179719.
XX
PA (FUSO ) FUSO PHARM IND LTD.
PA (OHNO/) OHNO T.
XX
PI Eda S, Matsuhisa A, Ohno T, Uehara H;
XX
DR WPI; 1994-035086/04.
XX
PT Probe for identifying bacteria causing infectious disease -
PT consists of a DNA fragment obtained by HindIII cleavage of the
PT pathogenic bacterial genomic DNA
XX
PS Claim 7; Page 62-64; 100pp; Japanese.
XX
CC The nucleotide sequence of a 3615 bp probe obtained by digestion of
CC Escherichia coli genomic DNA with the restriction enzyme HindIII.
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
CC by HindIII digestion of the genomes of Staphylococcus aureus,
CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
CC to detect their respective microorganisms in clinical samples.
XX
SQ Sequence 3615 BP; 966 A; 962 C; 805 G; 882 T; 0 other;

Query Match 7.1%; Score 108; DB 15; Length 3615;
Best Local Similarity 46.2%; Pred. No. 2.8e-21;
Matches 539; Conservative 0; Mismatches 590; Indels 37; Gaps 4;

Qy 312 tcactccagtgctgactgtgtctcaccggtatcccttaccattcagccattggccag 371
Db 2495 TCACACCATTCGTATGCTGATGGTTATCACACCCGTCACCTTCTGCTGGTGGGCGC 2436
Qy 372 caatgcgtgggtggcgatgtgtggcacacggtctacagggaacttatgattcgg 431
Db 2435 TATCAACCTGGTAAGCGAACTGATTCGCGCGGTTATCTCTGCGCTTATCAGCGGTC 2376
Qy 432 gtccagtcggcggtgtgtctctcgtgtgtgtgtactaccacatcgtcatcactgtc 491
Db 2375 CTGATTTGGGGCGGGTAATGGGGCGCTCTGGCAAAATCTGTCATGTTGGACTGC 2316
Qy 492 accagtcctcccccattgagctgtgttaccagggttgatcccttcattctcg 551
Db 2315 ACTGGGCGCTGGTCCGCTGTGTATCAATAACTTCACCGTGTGCGGTACGACACCATGA 2256
Qy 552 caacggcatctatg---gctaatacgcacagggtgcggcatgtttgagcagttcttc 608
```

RESULT 12

AAV52163/c

ID AAV52163 standard; DNA; 9769 BP.

XX

AAV52163;

XX

DT 23-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae genome fragment SEQ ID NO:30.

[illegible]





XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease; ds.  
 XX OS  
 XX Listeria monocytogenes.  
 XX WO200177335-A2.  
 XX PN  
 XX 18-OCT-2001.  
 XX PD  
 XX 11-APR-2001; 2001WO-FR01118.  
 XX PF  
 XX 11-APR-2000; 2000FR-0004629.  
 XX PR  
 XX (INSP ) INST PASTEUR.  
 XX PA  
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kroft J, Kuhn M, Ng E, Vazquez-Boland JA,  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX DR  
 XX WPI; 2002-010914/01.  
 XX XX  
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 XX  
 PS Claim 1; SEQ ID No 1; 192pp; French.  
 XX  
 CC The present sequence is the genome sequence of *Listeria monocytogenes*  
 CC EGD-e. This sequence and fragments of this sequence are useful for  
 CC selecting probes and primers for detecting genes in *L. monocytogenes* and  
 CC related organisms, and to study genetic polymorphisms and other genomes.  
 CC Proteins (AB847297-AB850149) expressed from the present sequence are  
 CC useful for raising specific antibodies, identification of *L.*  
 CC *Listeria monocytogenes* and related organisms, and for biosynthesis and  
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
 CC proteins encoded by it are also useful for selecting compounds that  
 CC regulate gene expression and cell replication and modulate *L.*  
 CC *Listeria monocytogenes*-related diseases. In addition, this sequence and proteins  
 CC encoded by it are useful in pharmaceutical and vaccine compositions for  
 CC the treatment or prevention of infections by *L. monocytogenes* and related  
 CC organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2944528 BP: 914202 A; 563301 C; 555061 G; 911964 T; 0 other;  
 Query Match 5.3%; Score 81.6; DB 24; Length 2944528;  
 Best Local Similarity 46.1%; Pred. No. 7.6e-12;  
 Matches 310; Conservative 0; Mismatches 359; Indels 3; Gaps 1;  
 Qy 171 caatgtggtccctgttgggttagatgtgcccagcggtaccagggccaccgtgttc 230  
 Db 2855687 CAACGCCACACTTATGCGGTACCAGTAACCTCCAGTACGTATGCTTCTGTTATTC 2855628  
 Qy 231 ctgtgctggtggtttcttggattctggaacgagtcgagaagttctctgcacagcgactca 290  
 Db 2855627 CGATTTTACTTGCAATTTGGATTATGCTTATGTGGAAAAATGGTGGATCGGATTATTC 2855568  
 Qy 291 agggcaactgcagactctctactccagtgctgagctgtgctctaccggattcccta 350  
 Db 2855567 CAACCTCGTTAAATATGCTGTTTGTCCCATTAATACAAATGTTATCGTTGTTCCCTTTGA 2855508  
 Qy 351 catctatgcgcattggcccgcaatgcgctgggtggcgagtgctggtcacacggtctac 410  
 Db 2855507 CGCTTGTGTTATTGGACCACTAGGTACATTTGTTGGAGACGGTGTATCTCGCGGAATTA 2855448

Qy 411 agggactttatgatttcggtggtccagtcggtggtcgtctctctctcgttgcgtgtgtactac 470  
 Db 2855447 ACTGTTTATTAATAAATAGGTGGAGCTATTGTTGGTATCTTGTGTTGGTGGCCATGGCCA 2855388  
 Qy 471 caatgctcatcactggtctgcaccagtccttcccgccaattgagctggagctgtttaacc 530  
 Db 2855387 TTATCGTTATGACTGGTATGCACATATGTTCCATTTGTTTATTTAGTAATTTAGCAA 2855328  
 Qy 531 aggggtggatccttcatcttcgcaacggcatc---tatggcacaatcgccacgggtgcgg 587  
 Db 2855327 AATATGGTTATGATAAATTTTACCGTTAACTTATATTTCCATATAGCCACACGGGAG 2855268  
 Qy 588 catgtttggcagtggttctctcctggcgaagagtgaagaagctcaaggccttcgagtgctt 647  
 Db 2855267 CGACTTTTCGGTGTATTTTTCGAGCAAAAGATAAAAACTGAAATCATTAGCATTTTCAA 2855208  
 Qy 648 caggtgtctcgcgtgttcttggattacaggagcctgcgatatctcgtgtgtaaccttcgcc 707  
 Db 2855207 CTGTTTAACTGTCATTAATGGGCGTAAACGGAACACGCGATGTACGGGATAAAACGTGTAT 2855148  
 Qy 708 tgcgtggtccgttctctcctcgtatcggtacgcagctatcggtggtgctgtgattgcac 767  
 Db 2855147 ATAAACGTCCTTTTCATGGCATCGTTAATTTGTTGGGCTGCTGGTGGCGGCTTTGCAATGA 2855088  
 Qy 768 tcttaatatcaaggcagtggtggtggcgctgcaggtttcttctggtgtgtttctattg 827  
 Db 2855087 TGTTGGTGTAAAGCGTATGTCCTTAACGTAAACGGTGGTATTCCAGGACTTCCAGGTC 2855028  
 Qy 828 atgtccagata 839  
 Db 2855027 TTGTCCGAGATA 2855016

Search completed: September 25, 2002, 09:14:10  
 Job time: 10631 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:05:49 ; Search time 2033.44 Seconds  
(without alignments)  
15714.662 Million cell updates/sec

Title: US-09-604-231-1

Perfect score: 1527

Sequence: 1 ctcattgcatctgcgcgctt.....gttgaacaccttgatgttcg 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hlg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Query Score | Match Length | DB ID | Description |
|------------|-------------|--------------|-------|-------------|
| -----      |             |              |       |             |

|   |    |        |       |        |   |            |                     |
|---|----|--------|-------|--------|---|------------|---------------------|
| c | 1  | 1527   | 100.0 | 1527   | 6 | AX069134   | AX069134 Sequence   |
|   | 2  | 1515   | 99.2  | 34980  | 6 | AX127152   | AX127152 Sequence   |
|   | 3  | 1492   | 97.7  | 1983   | 6 | AX122988   | AX122988 Sequence   |
|   | 4  | 1107.4 | 72.5  | 1109   | 6 | AX069136   | AX069136 Sequence   |
| c | 5  | 287.6  | 18.8  | 21838  | 1 | PDCRAFOPER | L32093 Pediococcus  |
| c | 6  | 287.6  | 18.8  | 21839  | 1 | PPSURFOP   | Z32771 P. pentosace |
| c | 7  | 253.4  | 16.6  | 2788   | 1 | AF401046   | AF401046 Lactobacil |
| c | 8  | 250.8  | 16.4  | 5800   | 1 | LL297015   | Z97015 Lactococcus  |
| c | 9  | 231.8  | 15.2  | 10264  | 1 | AE004395   | AE004395 Vibrio ch  |
| c | 10 | 230.8  | 15.1  | 2508   | 1 | STRSCRA    | M2711 Streptococ    |
| c | 11 | 218.6  | 14.3  | 10085  | 1 | AE006222   | AE006222 Pasteurel  |
| c | 12 | 215.2  | 14.1  | 2955   | 1 | VIBSCRAK   | M76768 Vibrio algi  |
| c | 13 | 210.6  | 13.8  | 2655   | 1 | SXSCRA     | X69800 S. xylosum s |
| c | 14 | 205.4  | 13.5  | 4158   | 1 | BSTREAPR   | Z54245 B. subtilis  |
| c | 15 | 201.8  | 13.2  | 22197  | 1 | D83967     | D83967 Bacillus su  |
| c | 16 | 201.8  | 13.2  | 37900  | 1 | D86417     | D86417 Bacillus su  |
| c | 17 | 201.8  | 13.2  | 208430 | 1 | BSUB0005   | Z99108 Bacillus su  |
| c | 18 | 194.8  | 12.8  | 341350 | 1 | AP003365   | AP003365 Staphyloc  |
| c | 19 | 190.6  | 12.5  | 2913   | 1 | AF269422   | AF269422 Staphyloc  |
| c | 20 | 190.6  | 12.5  | 2913   | 6 | AX144742   | AX144742 Sequence   |
| c | 21 | 183    | 12.0  | 299850 | 1 | AP001514   | AP001514 Bacillus   |
| c | 22 | 174    | 11.4  | 215050 | 1 | AL646084   | AL646084 Ralstonia  |
| c | 23 | 157.6  | 10.3  | 4487   | 1 | AF229829   | AF229829 Pseudomon  |
| c | 24 | 146.8  | 9.6   | 292550 | 1 | AP001513   | AP001513 Bacillus   |
| c | 25 | 146.4  | 9.6   | 296950 | 1 | AP001508   | AP001508 Bacillus   |
| c | 26 | 143.6  | 9.4   | 4621   | 1 | KPSCRYAB   | X57401 K.pneumonia  |
| c | 27 | 141.8  | 9.3   | 6950   | 1 | EWB250722  | EWB250722 Erwinia a |
| c | 28 | 138    | 9.0   | 5065   | 1 | ERWBGPA    | M81772 Erwinia chr  |
| c | 29 | 136.2  | 8.9   | 1398   | 1 | BSU34876   | U34876 Bacillus st  |
| c | 30 | 135.2  | 8.9   | 2819   | 1 | BACISPO    | D37921 Alkalophilol |
| c | 31 | 134.6  | 8.8   | 4322   | 1 | STSCRCOMP  | D67750 S. typhimuri |
| c | 32 | 134.6  | 8.8   | 231450 | 1 | AL596163   | AL596163 Listeria   |
| c | 33 | 132.2  | 8.7   | 3518   | 1 | BACSACP    | X73006 B. subtilis  |
| c | 34 | 132.2  | 8.7   | 97015  | 1 | BSENGR     | X73124 B. subtilis  |
| c | 35 | 132.2  | 8.7   | 200050 | 1 | AL591973   | AL591973 Listeria   |
| c | 36 | 132.2  | 8.7   | 212150 | 1 | BSUB0020   | Z99123 Bacillus su  |
| c | 37 | 130    | 8.5   | 1530   | 1 | STSCRA     | Y00541 Salmonella   |
| c | 38 | 126.6  | 8.3   | 3895   | 6 | BD003881   | BD003881 Polynucle  |
| c | 39 | 126.6  | 8.3   | 13589  | 1 | AE008535   | AE008535 Streptoco  |
| c | 40 | 126.6  | 8.3   | 14515  | 1 | AE007479   | AE007479 Streptoco  |
| c | 41 | 123.4  | 8.1   | 151947 | 2 | SPNEU1902  | AL449924 Streptoco  |
| c | 42 | 120.8  | 7.9   | 8006   | 1 | AF206272   | AF206272 Streptoco  |
| c | 43 | 117.8  | 7.7   | 5600   | 1 | AF205034   | AF205034 Clostridi  |
| c | 44 | 117.8  | 7.7   | 10611  | 1 | AE007557   | AE007557 Clostridi  |
| c | 45 | 114.2  | 7.5   | 136254 | 1 | ECOUW82    | L10328 E. coli; th  |

## ALIGNMENTS

|            |            |  |                          |     |        |                 |
|------------|------------|--|--------------------------|-----|--------|-----------------|
| RESULT     | 1          | AX069134   | 1527 bp                  | DNA | linear | PAT 25-JAN-2001 |
| LOCUS      | AX069134   | Sequence   | 1 from Patent WO0102583. |     |        |                 |
| DEFINITION | AX069134   | Sequence   | 1 from Patent WO0102583. |     |        |                 |
| ACCESSION  | AX069134   | Sequence   | 1 from Patent WO0102583. |     |        |                 |
| VERSION    | AX069134.1 | GI:12579016  |                          |     |        |                 |
| KEYWORDS   |            | Corynebacterium glutamicum.  |                          |     |        |                 |
| SOURCE     |            | Corynebacterium glutamicum   |                          |     |        |                 |
| ORGANISM   |            | Corynebacterium glutamicum   |                          |     |        |                 |
| REFERENCE  |            | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;              |                          |     |        |                 |
| AUTHORS    |            | Actinomycetales; Corynebacterineae; Corynebacteriaceae;              |                          |     |        |                 |
| TITLE      |            | Corynebacterium  |                          |     |        |                 |
| JOURNAL    |            | 1 (bases 1 to 1527)  |                          |     |        |                 |
| FEATURES   |            | Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberer, G. |                          |     |        |                 |
| source     |            | Orynebacterium glutamicum genes encoding phosphoenolpyruvat e;       |                          |     |        |                 |
|            |            | sugar phosphotransferase system proteins                             |                          |     |        |                 |
|            |            | Patent: WO 0102583-A 1 11-JAN-2001;                                  |                          |     |        |                 |
|            |            | BASF AKTIENGESellschaft (DE)   |                          |     |        |                 |
|            |            | Location/Qualifiers  |                          |     |        |                 |
|            |            | 1. .1527   |                          |     |        |                 |
|            |            | /organism="Corynebacterium glutamicum"                               |                          |     |        |                 |
|            |            | /db_xref="taxon:1718"  |                          |     |        |                 |
|            |            | 101. .1507   |                          |     |        |                 |
| CDS        |            |  |                          |     |        |                 |



/note="RXS00315"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC27276.1"  
/db\_xref="GI:12579017"  
/translation="MAWVPSLVNGYDVATMAAGEMPWMSLFGLDVVAOAGYGVLP  
VLVYSLIAETIEFLHKLRLKGLADFLITPVLTLLGLFLFIAIGPAMRWVGVLAHG  
LQGLYDFGGVPLGLLGLVSPITVITGLHQSPFPIELEFNQGGSTFFATASMANIAQ  
GAALAVFLFAKSEKLGAGSVAIVGTEPAIFGNLRLRWPFFIGTGAAGG  
ALTALENIKAVLAGAGFLGVSIDAPDMVFLVCAVVFIAFGAAIAYGLYLVRRN  
GSDIDPATAPVAGTIIKAEPAPFSDNSTIIQAPLGEAJALSSVSDAMPASGL  
GSGVAIVPRKQLVSPVSGKIVVAFPSGHAFVTRKEDGSNDYDILMHIGFDVNLNG  
THRNPLKKQDEKAGELCEFDIDAIAKAGYEVTPPIVVSNNKKIGPVNINYLGEIE  
AGANLNVAKKEAVPATP"

BASE COUNT 304 a 392 c 430 g 401 t

Query Match 100.0%; Score 1527; DB 6; Length 1527;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cteatgcatctgcgcgttcgcttcttccagtggttggttttccacgcgaaccaag 60  
DB 1 CTATGGCACTCGCCGCTTCGCCGCTTCGCGATGTTGGTTTCCACGCAACCAAG 60  
QY 61 cgttcgcgcgaatagttctggcgccgcttatgttgatggcgatggtgttcccgag 120  
DB 61 CGTTTCGGCGCAATGATTCCTGGCGCCGCGATTGGTATGCGGATGGTGTCCCGAG 120  
QY 121 cttggtgaacggttacgaactggtccgacacatggtcgcgggggaataatggttc 180  
DB 121 CTTGGTGAACGGCTACGACGTGGCGCCGACCATGGCTCGCGGGCAATGCAATGTGCTC 180  
QY 181 cctgttggttagatattgccaagcgggttaccaagggaacggtgtcttcgtgctggt 240  
DB 181 CCTGTTGGTGTAGATTGCCCCAAGCGGTTACAGGGGACCGGTCTCTGTGCTGGT 240  
QY 241 ggttcttgattctggcaacgactcgagaagttctctgcacaaagcactcaagggaactgc 300  
DB 241 GGTTCCTGGATTCTGGCAAGCATCGAGAAGTTCCTGCACAAAGCGACTCAAGGSCACTGC 300  
QY 301 agacttctgatactcactcagtgctgacgttgctgctgctcaccggattcttcaatcactgc 360  
DB 301 AGACTTCTGATCATCTCCAGTGTGCTGAGTGTGCTGCTCACCAGGATTCCTTACATTCATCGC 360  
QY 361 cattgcccgaacatgcctggtggtggtgctgctgctgctgctgctgctgctgctgctgct 420  
DB 361 CATTGGCCCAAGCATGCGTGGGTGGGGGATGCTGCTGGCACACCGGTCTACAGGGACATTTA 420  
QY 421 tgattcgggtggtccagtcggtgctgctgctgctgctgctgctgctgctgctgctgctgct 480  
DB 421 TGATTTCCGTTGGTCCAGTCGGCGGTCTGCTCTCGGTCTGGTCTACCTACCAATCGTCAT 480  
QY 481 cactggtctgcacagtccttcocgccaattgagctgagctggttttaaccagggtggtatc 540  
DB 481 CACTGGTCTGCACAGTCTTCCCGCCAATTTAGCTGGAGCTGTTTAAACAGGGTGGATC 540  
QY 541 cttcatcttcgcaacgcatctatgctgataatcgccagggtgcggcatgttttggcagt 600  
DB 541 CTTTCATCTTCGCAACGGCATCTATGGCTATATATCGCCAGGGTGGCGATGTTTGGCAGT 600  
QY 601 gttctcctgcggaagtgtaaaagctcaaggcgcttcgaggtgcttcagggtgcttcctgcgc 660  
DB 601 GTTCTTCTGGCGAAGAGTGAAGAGCTCAAGGGCCTTTCAGGTGCTTCAGGTGCTCCGC 660  
QY 661 tgttcttggtattacgagcctgcgacttcctggtgtaaaccttcgctgcgctggttcggtt 720  
DB 661 TGTTCTTGGTATTACGGAGCGCTGCGATCTTCGGTGTGAACCTTCGCCCTGGCGTGGCGGTT 720  
QY 721 cttcatcggtatcggtaccgcatctatgctggtggtggttgaattgacacttttaatacaa 780  
DB 721 CTTTCATCGGTATCGGTACCGAGCATCTCGGTGGCGCTTGTGATTCATCTCTTAATATCAA 780

QY 781 ggcagttgcttggtggcgctgcaggtttcttgggtgtgtttcttattgattgctccagatat 840  
DB 781 GGCAGTTGCGTGGCGCGCTGCAGGTTCTTGGGTGTTCTTTCTATTGATGCTCCAGATAT 840  
QY 841 ggtcatgttcttggtgtgtgcagttgttaccttcttcattcgcagcattcgcgcagcagattgc 900  
DB 841 GGTGATGTTCTTGGTGTGTGCAGTTGTGTACCTTCTTCATCGCATTCGCGCGAGGATTCG 900  
QY 901 ttatggccttacttggttgcgcgaacggtgcagcatgatccagatgcaacgctgcctcc 960  
DB 901 TTATGGCCTTACTTGGTTCGCCGCAACGGCAGCATTTGATCCAGATGCAACCGCTGTCTCC 960  
QY 961 agtcctgcgaggaacccaaagccgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 1020  
DB 961 AGTCCTTCGAGGACGACCAAGCCGAAGCAGAGCAGACCCCGCAGAAATTTTCAACGATTC 1020  
QY 1021 caccatcatcaggcaccctttgacgggtgaagctatgactgagcagcagcagcagcagcagcagc 1080  
DB 1021 CACCATCATCCAGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCACGCGATGC 1080  
QY 1081 catgttgccagcggaagccttgcgcggttgccgctgctccacccaagggcgagtt 1140  
DB 1081 CATGTTTGCACGCGGAAGCTTGGCTCGGGCTTGCCTATCGCTCCCAACCAAGGGCGAGTT 1140  
QY 1141 agtttccggtgagtggaagattggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1200  
DB 1141 AGTTTCTCGGTGAGTGGAAAGATTGTTGGTGGCATTCCTCATCTGCCCATGCTTTCGCAGT 1200  
QY 1201 tcgaccaaagcgtgagtggttccaatgttgatgctgctgctgctgctgctgctgctgctgctgctgct 1260  
DB 1201 TCGCACCAAGGCTCAGGATGTTCCAAATGTGGATATCTTGATGCAATTTGTTTCGACAC 1260  
QY 1261 agtaaacctcaacgcgcgcactttaaccgctgaagcgaagcgaagcgaagcgaagcgaagcgaagc 1320  
DB 1261 AGTAACCTCAACGCGCAGCCTTTAAACCCGCTGAAGAAGCAGGGCGGATGAAGTCAAAAGC 1320  
QY 1321 agggagcgtgctggtgaattcgtattgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1380  
DB 1321 AGGGAGCTGCTGTTGTAATTCGATATTGATGCCATTAAGGCTCAGGCTTATGAGGTAAAC 1380  
QY 1381 caegccagtggttcttcgaattacaagaaacccgacctgttaaacacttacggtttggg 1440  
DB 1381 CACGCCGATTTGTTTTCGAATTTACAAAGAACCGGACCTGTAACACACTTACGTTTGGG 1440  
QY 1441 cgaattgaacgagcagcgaacacctgctcaacgtcgcaagaaagcgaagcgaagcgaagcgaagc 1500  
DB 1441 CGAATTGAACGGGAGGAGCCACCTGCTCAACGTCGAAAGAAAGAGCGGTGCGCAGCAAC 1500  
QY 1501 accataagttgaaaccttgagtggttcg 1527  
DB 1501 ACCATAAGTTGAAACCTTGAGTGTTCG 1527  
RESULT 2  
AX127152/c  
LOCUS AX127152 349980 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 7068 from Patent EP1108790.  
ACCESSION AX127152 AX114121  
VERSION AX127152.1 GI:14041140  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 349980)  
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,  
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7068 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES Location/Qualifiers



```
source 1. 34980
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
/note="Seq 1 to long (3,309,400) split in 11, seq 7068
2,700,001 3,049,980"
BASE COUNT 81250 a 97718 c 90621 g 80391 t
ORIGIN

Query Match 99.2%; Score 1515; DB 6; Length 349980;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1526; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ctcaggcatcgcacctcgccttcgtcttcgacgttgccaggttggttggtttcaaccgaaccaag 60
Db 111326 ctcattggatcgtgcgcggttcggttcgttcgacgagtttggttggtttcaccgcaaccaag 111267
Qy 61 ggtttcggcgaatgagttcctggcgcgcgctgattgtagtgagtggttggtttcccccag 120
Db 111266 cgtttcggcgaatgagttcctggcgcgcgctgattgtagtgagtggttggtttcccccag 111208
Qy 121 cttggtgaacgctacgacgtggtgcgcgcacacatggtcggcggaatgcgaatgtggtc 180
Db 111207 cttggtgaacgctacgacgtggtgcgcgcacacatggtcggcggaatgcgaatgtggtc 111148
Qy 181 cctgtgttgtagtctgacgtggtgcgcgcacacatggtcggcggaatgcgaatgtggtc 240
Db 111147 cctgtgttgtagtctgacgtggtgcgcgcacacatggtcggcggaatgcgaatgtggtc 111088
Qy 241 ggtttcgttgatcttggaacgacgtgagagttcctgcacaagcgaactcaaggcactgc 300
Db 111087 ggtttcgttgatcttggaacgacgtgagagttcctgcacaagcgaactcaaggcactgc 111028
Qy 301 agacttcctgatacctccagtgctgacgttgctgtcaccggatccttaccatcctgc 360
Db 111027 agacttcctgatacctccagtgctgacgttgctgtcaccggatccttaccatcctgc 110968
Qy 361 cattggccagcaatgcgtgggtgaggtgaggtgctgtgcacacgctctacaggacttta 420
Db 110967 cattggccagcaatgcgtgggtgaggtgaggtgctgtgcacacgctctacaggacttta 110908
Qy 421 tgattcgtgtgctcagtcggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 480
Db 110907 tgattcgtgtgctcagtcggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 110848
Qy 481 cactggtctgaccagtcctcctcccaattgagctgagctggtttaaccagggtggtc 540
Db 110847 cactggtctgaccagtcctcctcccaattgagctgagctggtttaaccagggtggtc 110788
Qy 541 cttcattcgcacagcgcacatcgtggttaataatcgcacaggtgcggcgtggttggtcgt 600
Db 110787 cttcattcgcacagcgcacatcgtggttaataatcgcacaggtgcggcgtggttggtcgt 110728
Qy 601 gttctcctcggcgaagtggaagctcaaggccttcgagtgcttcaggtgctccgc 660
Db 110727 gttctcctcggcgaagtggaagctcaaggccttcgagtgcttcaggtgctccgc 110668
Qy 661 tgttttggtattacggagcctgcgacgttcgttggtggaaccttcgcctgcgtggtcgtt 720
Db 110667 tgttttggtattacggagcctgcgacgttcgttggtggaaccttcgcctgcgtggtcgtt 110608
Qy 721 cttcattcgcacagcgcacatcgtggtggtggtggtggtggtggtggtggtggtggtggt 780
Db 110607 cttcattcgcacagcgcacatcgtggtggtggtggtggtggtggtggtggtggtggtggt 110548
Qy 781 ggcaggtggttggcgtgcaggttctcgtggtggtggtggtggtggtggtggtggtggtggt 840
Db 110547 ggcaggtggttggcgtgcaggttctcgtggtggtggtggtggtggtggtggtggtggtggt 110488
Qy 841 ggtcattggttgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 900
Db 110487 ggtcattggttgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 110428

Qy 901 ttatggcctttacttggttcgcgcgaacggcagcagcttgatccagatgcgaacgcgtgctcc 960
Db 110427 ttatggcctttacttggttcgcgcgaacggcagcagcttgatccagatgcgaacgcgtgctcc 110368
Qy 961 agtgcctgcaggaacacacaaagccgaagcagaagcaccgcgcagaaattttcaaacgatic 1020
Db 110367 agtgcctgcaggaacacacaaagccgaagcagaagcaccgcgcagaaattttcaaacgatic 110308
Qy 1021 caccatcatccaggcacctttgaccgggtgaagctattgactgagcagcagcgtcagcagatgc 1080
Db 110307 caccatcatccaggcacctttgaccgggtgaagctattgactgagcagcagcgtcagcagatgc 110248
Qy 1081 catgtttgcagcgaagcagcgttggtcgcgcgttgccatcgtcccaacaaagggcagtt 1140
Db 110247 catgtttgcagcgaagcagcgttggtcgcgcgttgccatcgtcccaacaaagggcagtt 110188
Qy 1141 agttctccggtgagtggaagatgtgtggtcattcccatctggccatgcttctgcaatc 1200
Db 110187 agttctccggtgagtggaagatgtgtggtcattcccatctggccatgcttctgcaatc 110128
Qy 1201 tcgcaccaaggctgaggtggttccaatgtgatatctctgacatcgttggtttgcacac 1260
Db 110127 tcgcaccaaggctgaggtggttccaatgtgatatctctgacatcgttggtttgcacac 110068
Qy 1261 agtaaacctcaacggcgcacactttaaccgcgtggaagcagcagggcagcagatgaaagc 1320
Db 110067 agtaaacctcaacggcgcacactttaaccgcgtggaagcagcagggcagcagatgaaagc 110008
Qy 1321 agggagcgtgctgtggaattcgatattgattccattgaggtgaggttatgagtaac 1380
Db 110007 agggagcgtgctgtggaattcgatattgattccattgaggtgaggttatgagtaac 109948
Qy 1381 cagcgcgattgtgtttccgaattacaagaaacccggaccctgtaaacacttacggtttggg 1440
Db 109947 cagcgcgattgtgtttccgaattacaagaaacccggaccctgtaaacacttacggtttggg 109888
Qy 1441 cgaatttgaggcggagcacaactgctcactcgaacgcgaagaaagcagcagcgtccagcaac 1500
Db 109887 cgaatttgaggcggagcacaactgctcactcgaacgcgaagaaagcagcagcgtccagcaac 109828
Qy 1501 accataagttgaacacctgagtggtc 1527
Db 109827 accataagttgaacacctgagtggtc 109801

RESULT 3
LOCUS AX122988
DEFINITION Sequence 2904 from Patent EP1108790.
ACCESSION AX122988
VERSION AX122988.1 GI:14040476
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 1983)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 2904 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..1983
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 410 a 508 c 558 g 507 t
ORIGIN

Query Match 97.7%; Score 1492; DB 6; Length 1983;
Best Local Similarity 99.9%; Pred. No. 0;
```

|    |      |   |      |
|----|------|---|------|
| Qy | 1    | ctcatggcaatcgcgcggttcgcgttctcttgccagtggttggtttccacgcgcaaccaag   | 60   |
| Db | 481  | CTCATGGCATCTGGCGCGTTCCGCTTCTTGCCAGTGTGGTGGTTTCAACGCCAACCAAG     | 540  |
| Qy | 61   | cggttcggcggaatgagtctcctgggcgcgcgctgattggtatggcgatgggtgtcccgag   | 120  |
| Db | 541  | CGTTTCGGCGCAATGAGTTCTCTGGGCGCCG - GTAATGGTATGGCGATGGTGTTC       | 599  |
| Qy | 121  | cttggtgaacggctacgacgtggcgcgcaaccatggctgcgggcgaaatgcacaaatgtggtc | 180  |
| Db | 600  | CTTGGTCAACGGCTACGACGTGGCGCCGACCATGCTCGCGGCGAAATGCCAATGTGGTC     | 659  |
| Qy | 181  | ccgtgttggttagatgttgcccaagccggttacacagggcaaccgtgtctcctgtgctgggt  | 240  |
| Db | 660  | CCGTGTTGGTATTAGATTGTGCCAAGCCGTTTACAGGGCACCGTGTCTCTGTGCTGGT      | 719  |
| Qy | 241  | ggttcttgattcttgcaacgacatcgagaagtctctgcacaagcgactcaagggcactgc    | 300  |
| Db | 720  | GGTTCTTGGAATCTGGCAACGATCGAGAAGTTCTCTGCACAAGGCGACTCAAGGCGACTGC   | 779  |
| Qy | 301  | agacttctctgatactccagtgctgacggttgctgtctcaccggattctcttaccattcctgc | 360  |
| Db | 780  | AGACTTCTGATCACTCCAGTGTCTGACGTTGCTGCTCACCGGATCTCTTACATTCAATCGC   | 839  |
| Qy | 361  | cattggcccaagaatcgctgggtgggcgcatgctctggcacacagctctacagggaactta   | 420  |
| Db | 840  | CATTGGCCCAAGATGCGTGGGTGGGCGATGTGCTGGCACACGGTCTACAGGACTTTA       | 899  |
| Qy | 421  | tgatttcgggtggccagtcgcggcgtctgctcttcgggtctggtctactcaccaatcgctcat | 480  |
| Db | 900  | TGATTCGGTGGTCCAGTCGCGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTCAT    | 959  |
| Qy | 481  | cactggctgcacagctccittccgcgaattgagctggagcgtgttttaaccaagggtgcatc  | 540  |
| Db | 960  | CACGTGTCGTACACAGTCCCTTCGCGCAATTGAGCTGGAGCGCTGTTTAACCAAGGTGGATC  | 1019 |
| Qy | 541  | cttcatcttcgcaacggcatctatggctaaatgcgccaggggtcgcgcatgtttggcagt    | 600  |
| Db | 1020 | CTTCATCTTCGAACGGCATCTATGGCTAATATCGCCAGGGTGGCGCATGTTTGGCAGT      | 1079 |
| Qy | 601  | gttcttcctggcgaagagtgaanaagctcaaggccttcgaggtgcttcaggtgtctccgc    | 660  |
| Db | 1080 | GTTCTTCTCGCGAAGAGTGAANAAGCTCAAGGGGCTTTGCAAGTCTTCAGGTGTCTCCGC    | 1139 |
| Qy | 661  | tgttcttggtattacgagcctgcgactctcgggtgtaaccttcgctgcgctgagcgtt      | 720  |
| Db | 1140 | TGTTCTTGGTATTACGGAGGCTGGGATCTTCGGTGTGAACCTTCGCTGCCTGCCTGGCGTT   | 1199 |
| Qy | 721  | cttcaatgggtatcggttaccgcagcatcgtgtggcgctttgatgtcacctctttaatacaa  | 780  |
| Db | 1200 | CTTCATCGGTATCGGTACCCAGCATCTCGTGGCGCTTTGATTGACCTCTTTAATATCAA     | 1259 |
| Qy | 781  | ggcagttcggttgggcgcgcaggtttcttgggtgttttcttattatgctccagatat       | 840  |
| Db | 1260 | GGCAGTTGCGTGGGCGCTGCAGTTCTTGGGTGGTGTTCCTATTATGATGCTCCAGATAT     | 1319 |
| Qy | 841  | ggtcagtcttcgtgtgtcagttgttaaccttcttcacgtcattcggcgagcgattgc       | 900  |
| Db | 1320 | GGTCATCTTCTGCTGTGTGCAGTTGTTACTCTTTCATCGCATTCGCGCAGCGAATTGC      | 1379 |
| Qy | 901  | ttatggcctttaacttggttcgcgcgaacgcgacgatgatccagatatcaaccgctgctcc   | 960  |
| Db | 1380 | TTATGGCCTTTACTTGGTTCGCGCAACGGCAGCATTTGATCCAGATGCAACCGGTGCTCC    | 1439 |
| Qy | 961  | agtgctctgaggaacgaccaaaagccgaagcagaagcaccgcgagaattttcaaacgattc   | 1020 |
| Db | 1440 | AGTGCCTGCAGGAACGACCAAAAGCCGAAGCAGAAGCACCCGCGAGATTTTCAACGATTTC   | 1499 |
| Qy | 1021 | cacataatcaggcaacctttgacccgtgaagctatgactgagcagcgctcagcgatgc      | 1080 |
| Db | 1500 | CACCATCATCCAGCACCTTTTGCCGGTGAAGCTATTGCACTGAGCAGCGTCAAGCATGC     | 1559 |

Query Match 72.5%; Score 1107.4; DB 6; Length 1109;  
Best Local Similarity 99.9%; Pred. No. 1.3e-292;  
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 419 tatgatttcgggtggtccagtcggcggtctgctcttcggtctggtctactcaacaatcgtc 478
DB 1 TATGATTTTCGGCGGTCAGTCGGCGGTCTGCTTCCTGGTCTGCTACTACCAATCGTC 60

QY 479 atcaactggctgcacagtcctcccgccaattgagctgagctgtttaaccagggtgga 538
DB 61 ATCATGTGCTGCACCAAGTCCTTCCTCCGCAATGAGCTGGAGCTGTTTAAACCAAGGTGGA 120

QY 539 tccctcatcttcgcaacggcactctatggctaatatcgccaggggtgcggcatgtttggca 598
DB 121 TCTTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCAGGGTCGGCATGTTTGGA 180

QY 599 ggtctcttcctggcgaagaagtgaagaagctcaaggcccttgaggcttcacaggtctctcc 658
DB 181 GTGTTCTTCTCCGCGCAAGAGTGAAGAAGCTCAAGGGCCTTTCAGGTGCTTTCAGGTGCTCTCC 240

QY 659 gctgtcttggtattacgagcctgcgactcttcggtgtaaccttcgctgcgctggccg 718
DB 241 GCTGTCTTGTGATTAACGAGCCTCGGATCTTCGGTGTGAACCTTCGCCCTGCGCTGGCCG 300

QY 719 ttcttcacgtggtatcggtaccgcagctatcggtggtgctgtttgatgcactctttaatc 778
DB 301 TTTCTTCATCGGTATCGGTACCGCAGCATCGCTGCGCTTTCATGTCACCTCTTAAATATC 360

QY 779 aaggcagttgcgttggggcgcaggtttcttgggtgttgggtcttctattgattgctcagat 838
DB 361 AAGGCAGTTGCGTGGCGCTGCAGGTTCTTTGGGTGTTGTTTCTTATTTGATGCTCCAGAT 420

QY 839 atggtcatgtcttggtgtgctgagttgcttactcttcacgcttcacgcttcgacgagatt 898
DB 421 ATGTCATGTTCTTGTGTGTCAGTGTGTACCTCTTCATCGCATTCGGCGAGCGATT 480

QY 899 gcttatggcctttacttggctgcgcgaagcgcagcattgattccagatgcaacccgctgct 958
DB 481 GCTTATGGCTTTACTTGTGTCGCGCAACGCGCAGCATTTGATCCAGATGCAACCGCTGCT 540

QY 959 ccagtgctgcaggaaacacaaagcgaagcagaagcaccgcgagaaatttcaaacgat 1018
DB 541 CCAGTGCTTCGAGAACGACCAACGAAAGCCGAGCAGAACGCCGCGAGAAATTTTCAACGAT 600

QY 1019 tccacatcatcagcagcctttgacctgagctgagctattgctactgacgagctcagcat 1078
DB 601 TCACCATCATCCAGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGCTCAGCAT 660

QY 1079 gccatgtttgccagcggaagcttgctggcgtggcgttgccatgctcccaaccaggggcag 1138
DB 661 GCATGTTTGCAGCGGAAGCTTGGCTCGGCGCTTGCATGTCCTCCCAACCAAGGGCAG 720

QY 1139 ttagtcttcctcggtgagtggaagatttggtggcattcccatctcggccatgctttcgca 1198
DB 721 TTAGTCTTCTCCGGTGAGTGGAAGATTTGTTGGATTTCCCATCTGGCCATGCTTTCCGA 780

QY 1199 gttccacacagctgagatggttccaatgtggaattctgatactgacacattggtttcagc 1258
DB 781 GTTCGACCAAGCCTGAGAGTGGTTCCAAATGTGGATATCTTATGCACATTTGTTTCGAC 840

QY 1259 acagtaaacctcaacggcagcgcactttaaccgcgtgaagaagcagggcgatgaagtcaaa 1318
DB 841 ACAGTAAACCTCAACGGCAGGCACATTTAACCGCTGAAGAACGAGGGCGATGAAGTCAAA 900

QY 1319 gcagggggagctgctgtgaaattcgatattgagccattaaagctgcaggttatgaggtta 1378
DB 901 GCAGGGGAGCTGCTGTGTGAATTCGATATTTGATGCAATTAAGGCTGCAGGTTATGAGTA 960

QY 1379 accacgcagattgttctgaattacaagaacacggacacctgtaaacacttaacggtttg 1438
DB 961 ACCACGCCGATGTTGTTTCGAATTTACAAGAAACCGGACCTGTAAACACTTACCGGTTG 1020

QY 1439 ggcgaattgaagcggggcgcaaacctgctcaacgctgcgaagaagaagcgggtgcagca 1498
```

```
DB 1021 GCGAAATTTGAAGCGGGGACCCAACTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCA 1080
QY 1499 acacataaagttgaaaccttgagtggtcg 1527
DB 1081 ACACCAATAAGTTCAAACCTTGAGTGTTCG 1109

RESULT 5
PCR/AFOPPER/c PCR/AFOPPER/c
LOCUS Pedicoccus pentosaceus raffinose operon genes. 21838 bp DNA linear BCT 23-MAY-1994
DEFINITION Pedicoccus pentosaceus L32093
ACCESSION L32093
VERSION L32093.1 GI:493181
KEYWORDS agar gene; aqa gene; aql gene; alpha-galactosidase;
alpha-glucosidase; fructokinase; insertion element; permease; rafp
gene; rafr gene; raffinose operon; regulatory protein; scrA gene;
scrB gene; scrR gene; scrR gene; sucrose-6-phosphate;
transport protein.
SOURCE Pedicoccus pentosaceus (strain PPE1.0) DNA; Insertion sequence
IS30 homolog (transposable element Insertion sequence IS30 homolog,
kingdom Prokaryotae) DNA; Pedicoccus pentosaceus (strain PPE1.0)
DNA; Insertion sequence IS3 homolog (transposable element Insertion
sequence IS3 homolog, kingdom Prokaryotae) DNA; and Pedicoccus
pentosaceus (strain PPE1.0) DNA.
ORGANISM Pedicoccus pentosaceus
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Pedicoccus.
REFERENCE 1 (bases 1 to 21838)
AUTHORS Leenhouts,K.J., Bolhuis,A.A., Kok,J.J. and Venema,G.G.
TITLE The sucrose and raffinose operons of Pedicoccus pentosaceus PPE1.0
JOURNAL Unpublished (1994)
COMMENT On May 25, 1994 this sequence version replaced gi:475106.
FEATURES
source
location/Qualifiers
1..21838
/organism="Pedicoccus pentosaceus"
/strain="PPE1.0"
/db_xref="taxon:1255"
487..510
/note="DR2; putative"
/rpt_type=direct
488..561
/note="IR1; putative"
/rpt_type=inverted
513..561
/note="DR1; putative"
/rpt_type=direct
2724..2760
/standard_name="terminator 4"
/feature="putative"
complement(2929..3864)
/gene="rafr"
complement(2929..3762)
/gene="rafr"
/function="regulation raffinose operon"
/feature="putative"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="AAA25562.1"
/db_xref="GI:475107"
/translation="MNGEYKTLANKSPESNVLFQEAQLPNYTKGNVRDYSVIHY
IQEKGTFPAANHHPTATVLKAGDIFILPKGTGCTFYQNDNDQPKYFWGFSAGIRLEAM
LSGSLAOKCYLRQVQNGHIYADSELYKVLHINSLINDVLLGLIYLFYDLRLRWY
PADATNKKVSTEQENLAVYIQENYGTGCTIMDLCHYINLSRSLYTLFTHANTSP
IKLLTKLRLEDAKQRLSTSNNSVQSIANNVGYKDSFTTSKAFKRYSGASPSYRKSIG
I"
complement(3837..3842)
/gene="rafr"
complement(3859..3864)
/gene="rafr"
3909..5918
/gene="rafr"
-10_signal
-35_signal
gene
```



|    |       |   |       |
|----|-------|---|-------|
| Qy | 64    | ttcggcggcaatagttcttcctggcggcgcgcgtattggtgatggcgtatggtgttccocagctt | 123   |
| Db | 15264 | TTcGGTgcGCAccCCTATcTGGGTgcTAcG - ATGGGcATGATcATcGTCTTAcCATcAcT    | 15206 |
| Qy | 124   | ggtgaacgcctacagcgtgcccgcaccatggctgcggcggaatgccaatggtgtccct        | 183   |
| Db | 15205 | AGTTAAcGGCTATAcGGCTTGAACGACCATGGcAGCGGcGAAGATGTTAcTGGAAcGT        | 15146 |
| Qy | 184   | gtttggtttagatgtgccaaagcggattaccaggggcacogtgccttcctgtgctgggtggt    | 243   |
| Db | 15145 | CTTTGGGTAcAcGTTGCACAAGCGGCTATcAAGGCGcAAGTGTCTCCAGcTCTTAGCGGT      | 15086 |
| Qy | 244   | tcttgattcttgcaacgatcgaagaattctctgcacaagcgaactcaaggcgaactgcaga     | 303   |
| Db | 15085 | CGCCTTcATTCTAGcTAcCCTTGAAAAATTCTCCATAAAcAcATTAAGAGGGcATTGTA       | 15026 |
| Qy | 304   | cttctgaatcaactccagctgcagcttgctgcctcaacggattccttaccatcattcgcat     | 363   |
| Db | 15025 | CTTCAcGTTTAcCCCGATGTTGGcATTGTcATTAcTAcTGGTTTCTTAcACTTTTAAcATCGT   | 14966 |
| Qy | 364   | tggcccaacgaatgcgctgggtggcgatgtctggcgcacaggtctacagggactttatga      | 423   |
| Db | 14965 | TGGTCCCGTCTTAcGAcGGTGAcGATGcATTAAcTAAcGGTTTAcTAGGcTATATAA         | 14906 |
| Qy | 424   | tctcggtggtccagtcggcggtctgctctgttctggtctactcaacatcgctcatc          | 483   |
| Db | 14905 | CAGcACCGGCTGGATGTGTATGGGAATCTTTGGTTTATTATATTCTGcAATTGTTATTAC      | 14846 |
| Qy | 484   | tggctgcacacagtccttcccgcacaaattgagctggagctgtt-----taacca           | 531   |
| Db | 14845 | TGGTCTCCATCAAAcTCTCCAGCAATCGAAACCGcATTGTGGCCAAcTGTGCGCAAAc        | 14786 |
| Qy | 532   | gggtggatcttcattcttcgcaacgcgcatctatgctgaatatcgccaggtgcggcgatg      | 591   |
| Db | 14785 | TGGTGGTTCCTTTATCTCCcAGTCGGCTCGATGGCCAcACATGGGGAAGGTGCGCGAc        | 14726 |
| Qy | 592   | tttggcagtgcttctcctgcgcgaagatgaaaagctcaaggccttcagtgcttcaagg        | 651   |
| Db | 14725 | TTTAcGTATTCTTTTGCCAcTAAAGAcCAAAAGcAAAGCCCTGAcTCTTcCAGCGG          | 14666 |
| Qy | 652   | tgtctcgcgttcttgatttaacgagcctgcgattcttcggtgtgaaacttcgctcg          | 711   |
| Db | 14665 | GGTCTAcAGcCTAcTCCGGAATTAcGGAACCTGCTATTCTTCGGGGTCAAcCTCAAGATGA     | 14606 |
| Qy | 712   | ctggcgcgtctctcagctcgtgtaccgcagctatcggtggcgctatgattgacactctt       | 771   |
| Db | 14605 | ATTTCcATTGCTCTTGcAGCGcATTGGCTCAGGAATTGCTTAcGCTTTCTTAGAcATT        | 14546 |
| Qy | 772   | taatacaaggcagcttcggttggcgctgcaggtttcttgggtgtgtttctattgatgc        | 831   |
| Db | 14545 | CCAGcTTCTGTCGGTGGcATGGGTCCCGcATcAGTcATCGGAATTCTTCGATTCTTCGATTGCTC | 14486 |
| Qy | 832   | tceagatatggtcatgttcttgggtgtgtgcaagttgttaccttcttcacatcgccgcg       | 891   |
| Db | 14485 | AAAGTCGATTCCAcGATTcATGTcAGTGTGTcATCTCGTTCGTCTCGcATTTATTTC         | 14426 |
| Qy | 892   | agcgattgcttatg  | 907   |
| Db | 14425 | AACTTTTATCTATGCC  | 14410 |

RESULT 6

PPSURFOP/c

LOCUS PPSURFOP 21839 bp DNA linear BCT 24-MAY-1994

DEFINITION P.pentosaceus (PPEI.0) sucrose and raffinose operons.

ACCESSION 232771

VERSION 232771.1 GI:493728

KEYWORDS alpha-galactosidase; alpha-glucosidase; enzyme Iiabc; fructokinase; insertion element; permease; regulator; sucrose.

SOURCE Pediococcus pentosaceus.

ORGANISM Pediococcus pentosaceus

Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Pediococcus.

```

/db_xref="GI:475964"
/db_xref="SWISS-PROT:P43466"
/translation="MQEENHYKVGRLIYFGCAKGNDAFYSLISGLYLIIFITSHLFD
TGNKALDNRMVSLVTILIMVLRIVELFIDPFIAGNADTKNSGPHRPWVVGCTYSS
IILLFLFTNGLGLYAKNMLYLVVFAILYITMDIFVSKDVGFWMLPSLTWDSRERE
KNATFARLSTIGGLGLVGLVMPAVIPFSAKATSGDNRGWFIFALICLIALISAWG
VGLGTREVSDIRNKQDVTGVNEIFKALAKNDQLLWALAYLFYGVGINLGSLEVY
YFTYIMGRKPSISILINIFLGIATISLFPVLSKFKSRKGFAGCLVPMGLGIALFT
IAGSNLWLLAATMEFPFQOMVFLVLMVITDSVEYQGLKGRHDESIALSVRPLID
KFGGALSNGVGQIALISGMTTGATASSITAAQOLHFKLTMFAFPALMLLIAIGIFSK
QIPLTEKHAELVAELERWTWKFDNTDOVAEKVYTSIDLATPIAGOVIPLAQVNDP
TRAAGTLDGPAIKPSDGRILLAPFDVATVQVTTTRHAGLVGNDGIVLLIHGLGVTK
GGAVKATDNIQLHTKA"
gene
5976..8177
CDS
/gene="agar"
/gene="alpha-galactosidase"
/standard_name="alpha-galactosidase"
/citation=[1]
/codon_start=1
/transl_table=11
/evidence=experimental
/product="alpha-galactosidase"
/protein_id="CAA83665.1"
/db_xref="GI:475965"
/db_xref="SWISS-PROT:P43467"
/translation="MSLITVDOANRVFHLNHTLSYIFAVEOGGTLISLHYFGHVDHY
HGELRYPRVDGFSGLNPGSDTPTFSRDTLPKXYSTAGEMDYHLPAAVYRHVDGANL
YLYQYRIEAKGPKLSGLGPAAFVEDETEAEITLIVLDOVQSOVEFDLQYTIYDRPV
FTRSVQVCOGHAVNLEKVASMQIDFTDRQFETITPLGAHANERPERGSINYGITQ
FGSLRTSSHNPFLALVHTTTFESGDYGFNLVYSGNHAFELEKQDLQDLHLVVG
INSYNFNLKAGATFQTPLEVMVYTNKGLNAMSQAYHHLIRERVVRSFKNOERPIV
VNNWEATFVDNEAKLPIVDRAKOLGIEMLVDGCGFGRDDNSSLGDMQVDHKKF
ROGLNHVYVHEOGLKFGIMLEPEMISYDSKLYQHPDYLMQVGRSPSPRNQYIL
DUGRQARNNIFDQLQKSKQIDIKWDMNHLSDIYSVALPPRQGEVYHRYVIG
LYELLERLTAPHILFEGCGGGGFRDAGMAYMPOIWAUSDVARTLTQYGTSL
APYISATIRSVSPNQQTGRETSRSTSAVASAGYGLYELDLTQLSSAKQIVQKQV
VOYKQRLPIQFGEFVRLKSPITSNOAAMFVSPQODEAIVMVFNLTSVAQPSLTKTK
LVCLNPKLANQIATKAFEGDELMQLGFDYDPVYQDYTKVYHFKAVTEN"
8265..8289
/standard_name="terminator 3"
/citation=[1]
9403..9426
/notes="DR2"
/citation=[1]
/rpt_type=DIRECT
/complement(9519..10424)
/gene="iso-IS30"
/complement(9519..10424)
/partial
/gene="iso-IS30"
/standard_name="IS-element"
/function="IS-30 (E.coli) like IS element"
/citation=[1]
/codon_start=1
/transl_table=11
/product="insertion element"
/protein_id="CAA83666.1"
/db_xref="GI:475966"
/db_xref="SPRMBL:Q52261"
/translation="SDAVTRKSASLKWCLPOTHLGADLLSSIYSERIKIETFEEL
GLSNTQMGRVLRNRPSTISYELSRQPYQAEIAQTDAEYKRSQCGRKTKLSDELQKI
LNHLRLSWSPGMIAEHFKLATKSIYNLWLNQGRIFGSLNDLPEHGRQRRNVPQRSYN
QSLGRSIEORPMFNORNRIQDFELDTVVGPRGHSKAVLLTLIDRKSRLFWLRKDR
TWTATNEALTKELTFENGVPVHSFTVDRGTEFSGLSYLSQYGIKITYYCHAYTPADVW
MNALGIYVFIKGLVLSLVLKI"
10396..10442
/notes="DR1"
/citation=[1]
/rpt_type=DIRECT
10397..10443
/notes="IR2"

```

```

/terminator
/citation=[1]
/rpt_type=INVERTED
10681..10716
/standard_name="terminator 2"
/citation=[1]
/complement(10727..11593)
/gene="scrk"
/complement(10727..11593)
/gene="scrk"
/standard_name="fructokinase"
/function="fructose 6-phosphotransferase"
/citation=[1]
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="CAA83667.1"
/db_xref="GI:475967"
/db_xref="SWISS-PROT:P43468"
/translation="MLLGAIEAGGTFFVCATGAENGQVSDRISIPITTPVETMTAVDD
YFTTHPDAIGIGSFGPIGVNPHDPKYGITTTTPKPGWGFDFLGLHKSOFNPLVYMT
TDVNEAAYGESMIGIAKDPVNSIYMTIGTVGAGVISOHNIEPNRTHTELGHMRLNL
PGDDPKSNCPYHDICLEGLAAGPAVGKRTGKAGKDIPVDDPWPPIITDIYAQCYNLT
VAFAPDKIILNGVMNQRLQFFPMIREKFAAYLNGYEVEVPLDDYIIVPAGLNNNGIAG
GLLLAOALAKNA"
gene
/complement(13852..15989)
/gene="scrA"
/complement(13852..15807)
/gene="scrA"
/standard_name="enzyme IIsr"
/function="enzyme II of the PTS system, sucrose specific"
/citation=[1]
/codon_start=1
/transl_table=11
/evidence=experimental
/product="enzyme IIsr"
/protein_id="CAA83668.1"
/db_xref="GI:475968"
/db_xref="SWISS-PROT:P43470"
/translation="MNHQFVADRVLNAIGKNNIQAAHCAATRLRLVLIKDESKIDQOAL
DDADVKQGFETNGOYQIILIGPDVDKVYDALLVKTGLEVTDDDKAVAAAGQKNRP
LMDFLKLVSDIFIPVLPALVAGLLMALNNVLAERHFAKSVVEYVPGKLGIAEMIN
AMASAPFTFILLGSATKRFNGNPYLGATGMIMVPSLVNGYSVATMAAGRMVY
WNVGLHVAQAGYQGVLEVLGAFILATLEKFFHKGAFDFETPTMFAIVITGFL
TFTIVGPLRVTSDALTNGLVGLYNSGTWIGMIFGLLYSAIVITGLHOTFPAIETQL
Query Match 18.8%; Score 287.6; DB 1; Length 21839;
Best Local Similarity 58.8%; Pred. No. 1.9e-67;
Matches 539; Conservative 0; Mismatches 364; Indels 13; Gaps 2;
QY 4 atggcatctgcgcgtttcgttccagtggttgggtttccaccgcaaccaagcgt 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15324 ATGGCTAGTCGCGCGTTTACTTTCTTGCCAATCCTCTTAGGATTTTCAGCAACGACGC 15265
QY 64 ttccggcggaatgagttccctggcgcgcgatgttgatggcgatggtgttcccgactt 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15264 TTCGTGGCAACCCATCTCTGGGTGTACG-ATGGGCATCATGTGCTTACCATCACT 15206
QY 124 ggtgaacgctagacgtggtgcccaccatggtgctggcggaataccaatggtgcct 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15205 AGTTACGGCTATAGCGTTCACAGCCATGGCAGCGCGGCAAGATGTTTACTTGAACGT 15146
QY 184 gtttgggttagatgttgcccaagccggtttaccagggcacccgtcttctctgctggtgg 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15145 CTTTGGGTACACGTTGCACAGCCGGCTATCAAGGCCAAGTCTCCACACTCTTAGCGT 15086
QY 244 ttcttgattctggcaacgactgagagtgcttcgcacaagcgaactcaagggcactgcaga 303
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15085 CGCTTCATTTAGTACCTTCCAAAATTTCTTCCATAAACACATTAAGGGGCAATTGA 15026
QY 304 ctctcgatcactccagtgctgagcttgctgcacggattcttaccatccatcgcacat 363
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15025 CTTACGCTTACCCCGATGTTGCCATGTGATCTGTTTCTTAACCTTTTACAATCGT 14966

```





|    |      |   |      |
|----|------|---|------|
| QY | 482  | actggtgcacacagctcctcccgccaattgagctggagctgttt-----aac              | 529  |
| Db | 482  |   |      |
| QY | 622  | ACTGGCTTACATCAAAAGATTCCCGCAAATTGAAACGACACTTTTGGCAGATAATTGCCAAA    | 681  |
| Db | 622  |   |      |
| QY | 530  | cagggtggatcccttcacatctcgcaacgcgcatactatggctaataatcgcccagggtgcgcga | 589  |
| Db | 530  |   |      |
| QY | 682  | ACTGGTGGAATCGTTTTATTTTCOCGTGCAGCGAATGGCAATATTGCTCAAGGGGGCTGCA     | 741  |
| Db | 682  |   |      |
| QY | 590  | tgttggcagtgcttctctgcgcgaagagtgaagaagctcaaggcgccttgaggtgcttca      | 649  |
| Db | 590  |   |      |
| QY | 742  | ACTTTCGCTGTAATCTTCGTACTAAGAATAAACACAAGTCATTAACGACTCTCTGCT         | 801  |
| Db | 742  |   |      |
| QY | 650  | ggtgtctccgcgtgtctcttggtattacaagagccgcgcgatcttcggtggaaccttcgcgtg   | 709  |
| Db | 650  |   |      |
| QY | 802  | GGGATTTCTCGCATGTGTGGGAATTACTGAACGACGATTAATTGGGGTTAAATTTAAAAATTG   | 861  |
| Db | 802  |   |      |
| QY | 710  | cgcggcgtctctcatcggtatcggtaccgcagctatcgatggcgtcttgattgcactc        | 769  |
| Db | 710  |   |      |
| QY | 862  | AAGTTTCCAATCTTTTATTTGGTTTAAATTGCATCAGGAATCTCATCGTTTATTTATTTGGTTTA | 921  |
| Db | 862  |   |      |
| QY | 770  | ttaatatcaaggcaggtgcgtgggcgcgcgcaggtcttctgggtgttgttcttatgat        | 829  |
| Db | 770  |   |      |
| QY | 922  | TTACATGTTTTATCACATATCAATGGGACCTGCAGCAATTAATGGGTTAATTTCGATTTGCA    | 981  |
| Db | 922  |   |      |
| QY | 830  | gctccagatagdcagtgctctgtgtgtgtgcagttgttacctcttctaatgcattcgcgc      | 889  |
| Db | 830  |   |      |
| QY | 982  | CCTAAGAGCATCCCTAGTTTTTATGATGGGAGCTATTATTAGTTTCGTAATTCGCTTTGTG     | 1041 |
| Db | 982  |   |      |
| QY | 890  | gcagcgatigtctatggcctttacttggttcgcgcgaacgcagcattgatccagatgca       | 949  |
| Db | 890  |   |      |
| QY | 1042 | GGGACATATTATACGG-----fTAAAAAGGCA                                  | 1068 |
| Db | 1042 |   |      |
| QY | 950  | accgcgtctccagtgctgcaggaacgccaaagccgaagcagaagcacccgcagaattt        | 1009 |
| Db | 950  |   |      |
| QY | 1069 | ATGAAGACAACATGAAGAAGAAATAATCAATGAAGCACCAGCTACCCGACGAAGTAGTGGAG    | 1128 |
| Db | 1069 |   |      |
| QY | 1010 | tcaacgattccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagc       | 1069 |
| Db | 1010 |   |      |
| QY | 1129 | AGATTACAAGATGAAGAATTAATAGTCACAGCTTACCGGCAAAATTTGTACTTAGCATCA      | 1188 |
| Db | 1129 |   |      |
| QY | 1070 | gtcagcgatgccattgttcagcgcggaagcttgctgcgcgcgttgccatgcctcccaacc      | 1129 |
| Db | 1070 |   |      |
| QY | 1189 | GTACCTGTATCCAGTTTTTGCAAAGTAGAGCAATGGGAAAGGCAATTCGATTTATGCCAACT    | 1248 |
| Db | 1189 |   |      |
| QY | 1130 | auggggcagttagttctccgcgtgagtgaaaagattgtgtggcattccccattctggccat     | 1189 |
| Db | 1130 |   |      |
| QY | 1249 | TCTCAGGATGTACTTGCACCACTTACCG-----TGTGATAACAATTCGGCTAATACT         | 1302 |
| Db | 1249 |   |      |
| QY | 1190 | gcttctgcagatttcgaccaagcgtcagagatggttccaatgttgatatttgatgcacatt     | 1249 |
| Db | 1190 |   |      |
| QY | 1303 | GGTCAACCATACGGGATAAAATTCGGATGATGTGTCAGAAAGTG-----CTAATTCATATT     | 1356 |
| Db | 1303 |   |      |
| QY | 1250 | ggtttcgcacagtaaacctcaacggcgcgcactttaaccgcgtgaagaagcagggcgat       | 1309 |
| Db | 1250 |   |      |
| QY | 1357 | GGTTTAGATACAGTTAATTTAAATGGTATAGTTTTGAAAAGATGTCCCAACAGGGACAA       | 1416 |
| Db | 1357 |   |      |
| QY | 1310 | gaagtcaaacgaggaggagctgctgtgtaattcgatatattgatgccattaagcgtgcaggt    | 1369 |
| Db | 1310 |   |      |
| QY | 1417 | CATGTTAGGAAGCGCATTTATTAGTGCTATTTTATGATTTGATTAAGATTAAACAGCCGG      | 1476 |
| Db | 1417 |   |      |
| QY | 1370 | tatgaggttaaccacgcccgattgtgtttcgaat                                | 1402 |
| Db | 1370 |   |      |
| QY | 1477 | CTAACACCGCTAACCAATGACTATTCTGTGACGAAT                              | 1509 |
| Db | 1477 |   |      |

**RESULT 8**

|            |   |         |     |        |  |                 |
|------------|---|---------|-----|--------|--|-----------------|
| LL297015   |   |         |     |        |  | BCT 05-APR-1999 |
| LOCUS      | LL297015  | 5800 bp | DNA | linear |  |                 |
| DEFINITION | Lactococcus lactis cremoris sucrose gene cluster.   |         |     |        |  |                 |
| ACCESSION  | Z97015  |         |     |        |  |                 |
| VERSION    | Z97015.1  |         |     |        |  |                 |
| KEYWORDS   | enzyme II sucrose protein; fructokinase; sacA gene; sacB gene; sack gene; sacR gene; sucrose-6-phosphate hydrolase. |         |     |        |  |                 |



|                           |   |
|---------------------------|---|
| -10_signal                | complement(2737..2743)<br>/gene="sacA"<br>/evidence-experimental<br>complement(2761..2767)<br>/gene="sacA"<br>/evidence-experimental<br>2815..2821<br>/gene="sacB"<br>/evidence-experimental<br>2815..4835<br>/gene="sacB"<br>2837..2843<br>/gene="sacB"<br>/evidence-experimental<br>2872..2877<br>/gene="sacB"<br>2883..4835<br>/gene="sacB"<br>/EC_number="2.7.1.69"<br>/codon_start=1<br>/transl_table=11<br>/product="enzyme II sucrose protein"<br>/protein_id="CAB09690.1"<br>/db_xref="GI:4581477"<br>/translation="MNHQVAERILNAVGRDNTQGARHCATRLRLVLKDTGVDAQAL<br>DNDPDLKGTFEAAGQYQIVPGDQVNTVVEEFIKLTGISEASTADIKEIAGSQKQNP<br>VMAVLKLTGIFVPLIPALVAGGLLMALNALTAEHLFATKSLVEMFPMWKGFDADIV<br>TMSAATPMPILIGYSAYKRCFGNPYLGAVGMIWMPCLINGYNVAEAIASHHTWY<br>WDFGPKVAQGOQOVLPIGVAFILAKLERFPHKYLNDADIDFTPTLLSVLTGCLF<br>TFTIVGPALEFVSNGLTDGLVGLYNTLGALGLMVFYGSAYVVTGLHQSFPALFETML<br>ITNYHSGGDDIFVPAACANAQAQATFALLVTKNKTALAAAPAGVSAITLGITE<br>PALFGINKLKLKIPFFALGASAIISFLMGLFHVLAIVSLGSLIGISIKAGYNLQFM<br>ISFISFLAVTSIYGRRMEAKSITTEKKNQATQYQPKVIIDPVKSGELLAPI<br>NGVPIPSDVDPVFSKEIMKGIAIKPKSGELFSPADGEIITAVETGHAYGKTKNG<br>GEVLLHIGIDTVSMNGNPTONVKGQVKAGDLLCSFQKEEIKKSLDDTVIIVTIN<br>SASYNELILSENVDIKVEKILLN" |
| RBS                       | 4838..4842<br>/gene="sacK"<br>4838..5720<br>/gene="sacK"<br>4848..5720<br>/gene="sacK"<br>/EC_number="2.7.1.4"<br>/codon_start=1<br>/transl_table=11<br>/product="fructokinase"<br>/protein_id="CAB09691.1"<br>/db_xref="GI:4581478"<br>/translation="MSVYGSIEAGTGKTFVLAIADENFIIKKRFATTPQETISKT<br>IKYFKNRVSAYGLGSGFIDMLNSKTYGTTSTPKVGKNNINLVGOLKEALDIIPIY<br>FTDWNASAYGEMKNTGINKLVYLTFTIGTGGAIONGVFIGGIGHSEMGHORINRHR<br>DVLTFEGICPFHGDLEGVAAGPSLEARTGILGEKISSDDPJDILISVYLAQAANAT<br>LTLAPBCIILGGGVMEKPNWISLIQKFIISMLNNYIDLPCSVEKYIRLITVVKNGSAT<br>LGNFYLAISLFYKE"   |
| BASE COUNT                | 1858 a 915 c 1031 g 1996 t  |
| ORIGIN                    |   |
| Query Match               | 16.4%; Score 250.8; DB 1; Length 5800;  |
| Best Local Similarity     | 51.4%; Pred. No. 2e-57;   |
| Matches 725; Conservative | 0; Mismatches 647; Indels 46; Gaps 4;   |
| QY 4                      | atggcatctgcgcgttcggttcttccagtggtggtggttccacgcgaacaacgct 63<br>  |
| Db 3366                   | ATGAGTCCGCACCATTTACTTTATGCCAATTTTAATAGGTTATTCCTGCGCAAAAAGA 3425<br>   |
| QY 64                     | tccgcgcgaatgagttcttcggcgccgcgctattggtatgcgatggtgttcccgagctt 123<br>   |
| Db 3426                   | TTTGGCGAAATCCATATTAGGTGCTG-TTGTAGGTATGATTATGCTGCTGGACT 3484<br>   |
| QY 124                    | ggtgaacggtacagctggccgccacacatggtcgcggcggaatgccaatggtgctcct 183<br>  |
| Db 3485                   | TATTAAATGGATATAATGTTGCTGAAGCAATCTCTAATCATACATAGCTTATTGGGATAT 3544<br>   |

Qy 1246 cattggtttcagacagtaaacctcaacggcaccacacttttaaccgcctgaagaagcagggg 1305  
 Db 4598 TATTGGAATAGATCTGTTTCAATGAATGGTAATGGATTCTACAAAATGTTAAAGTTGG 4657  
 Qy 1306 cgatgaagctaaacagcagggagctgctgtgtgaattcgattgattgattccattaaggctgc 1365  
 Db 4658 CCAGAACTAAACAGCGGGATTACTAGGATCTTTTGATTAAGAAGAAATCAAGAAGAG 4717  
 Qy 1366 aggttatgagtgacacacgcccgattgtttgttcgaatt 1403  
 Db 4718 TGGATTGCATGATCTGTAATTAATGTTATTATCAAAATT 4755

RESULT 9  
 AE004395/c  
 LOCUS  
 DEFINITION Vibrio cholerae chromosome II, section 52 of 93 of the complete chromosome.  
 ACCESSION AE004395  
 VERSION AE004395.1  
 KEYWORDS GI:9658068  
 SOURCE  
 ORGANISM Vibrio cholerae.  
 Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
 DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
 Nature 406 (6795), 477-483 (2000)  
 JOURNAL  
 MEDLINE 20406833  
 REFERENCE 2 (bases 1 to 10264)  
 AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
 Direct Submission  
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source  
 1..10264  
 /organism="Vibrio cholerae"  
 /strain="N16961"  
 /serotype="O1"  
 /db\_xref="taxon:666"  
 /chromosome="II"  
 /note="biotype: El Tor"  
 complement(114)..1553)  
 /gene="VCA0653"  
 complement(114)..1553)  
 /gene="VCA0653"  
 /note="similar to GB:M76768 SP:P22825 PID:155262; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="PTS system, sucrose-specific IIBC component"  
 /protein\_id="AAF96554.1"  
 /db\_xref="GI:9658069"  
 /translation="MDYPVIAKOLLESGLGKSNIQALAHCAHTRLRVLNDETQINESA IESLQVGKQFVAGQYQIIFGSGIVNQVYAEAKLTGLVEMSTNDVASAEKQNNWA QRAVGLSDIFRIFAIIVAGGLMGLFNLVITAPLGFIEQSLIDANPGLADIASMIN TFANAPFVLPVLLAFSASRKFSGNPFGLAALGLMVLHPDLLNGWFGSGVSGTPT WNILFEIEKVGQGVSVLPVLSAYILAKTENGRLKMLVPSVDNLLTPMLAIFITGFL TFVVGPLRDVGMGLDNLWYDSAGFVGALFCFIYAPVITGMHHSFIAETQL LADIVTGTGTFEIPPIAMNSHAGAAALAVGYMTKETKLGVAIPSGVITALLITEPA MFGNKLRIPTFPAICGAALASAFITLFNVKAQAALGAAGLPGIISINPQQIGYING

gene  
 CDS  
 MATSVAAPALVTLLAMRETKQAQVTA"  
 1854..2846  
 /gene="VCA0654"  
 1854..2846  
 /gene="VCA0654"  
 /note="similar to GB:L14685 PID:289864; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sucrose operon repressor ScrR, putative"  
 /protein\_id="AAF96555.1"  
 /db\_xref="GI:9658070"  
 /translation="MASLHDVARLAGVSKSTVSRVINDEYGVKESTKIKVLKAVEQCG YMNQVAKDLKSQKTNLIGIVIPRVSSHATAQGVGLTAIFEQAGKRVLLANTHQVHA KEIYEIQFNQKRVGIVFYATHLDQPLVHAIOHSAVPVVLVGQDGLYIPSTIVHDD SIFSGYQMAQVLDYQPLDGLFCATDRLAVGALKALQELGVHVQGVKLLGCVNDEL AYVSNPLSTFNFAFDKAGENAAMLLERAGRGQEMSKVYLTFQNIQIRETCPV"  
 2922..4568  
 /gene="VCA0655"  
 2922..4568  
 /gene="VCA0655"  
 /note="similar to GB:M74035 PID:155269; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sucrose-6-phosphate dehydrogenase"  
 /protein\_id="AAF96556.1"  
 /db\_xref="GI:9658071"  
 /translation="MLDITLLELAGGINNVTRILAPQGVVLAKLHPPLVPHLPDDVS QSLVGEWQLSVQRTAEVSDQAAIGKAIAERQKLETLPYQTALDCPYRPLWHISPP LQGLNPNNGFIHQGEYHLEFYQWHPFACEHKRKYVHLKSLDLVDMQWQSVALTSPDW FSHGVFSGHAVSHOODLWLFYTNTRIGVDOROTMOCARNMNNEFEKLGPIVRC LPSEVTEHIRDPKVIYTOGKQMLGATLAHOGRLAVYHSDOLLHWHFQKLGDELG DYGMWECPDWFELOEAEAFVFGPGIATSNAPHTTEHQNRIFRATONAGEALALOG WPIDEGFDYAPQATADRGVLCVGMGLPDETQHPSCDQGWIHQITALELWREG RIYQHPRLRELDTSATQIRQGDITRELKDSPTVELRLADQSLEIFINOGEHVMTS RTTAPDASGISLUGASVDKLVYMAPASAPFNLEVNVP"  
 complement(4690)..5661)  
 /gene="VCA0656"  
 complement(4690)..5661)  
 /gene="VCA0656"  
 /note="similar to GB:M76768 SP:P22824 PID:155263; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="fructokinase"  
 /protein\_id="AAF96557.1"  
 /db\_xref="GI:9658072"  
 /translation="MNAVLRLSSNHLIRSDMSRVLTGDAVVLDLIPDGOHYLKCPG GAPANVAVAIARLCGRSAPFRVGNPDPCRFMOQTLTDEQVDCQHLFYDVRHRTSTVY VLDDEHGRSFTFWVPKPSADQFLQSDIPSFQNGEWLHVCSIALANQPSRSTFAAIA QMKVEGIVSFDPNLRNEWSEPELQATVRAVGLADVVFSEELQFTLTGTQSTIEE GLQAIADFOIPLVVTIGAKALVTPNSRQIVSGKAVKPIDITGAGDAFVGLLYRL SVAQDWHNQATILDAVKWANGCGALATTQKGAMTALPNAALYAFLE"  
 6008..7546  
 /gene="VCA0657"  
 6008..7546  
 /gene="VCA0657"  
 /note="similar to SP:P13035 GB:M21277 GB:M55989 GB:M96795 PID:146181; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="aerobic glycerol-3-phosphate dehydrogenase"  
 /protein\_id="AAF96558.1"  
 /db\_xref="GI:9658073"  
 /translation="MNASTRFTQDEVLDTIIVGGINGAGIAADATGRGLKGLYDA KDFASSTASSKLHVGRLHYEYFRLSEALAREVLLKPAHPHITPMRRLPHR PFLTPAWMIRAGILVYDNLGKRTSLPASHKLVNKAQSVTKPEMIGEEYSDCVDDAR LVLLNAMQEQBAEVLNCTVEKAEMGLDNLVLLDEQTCFGRFESHCWNAAGP WVKQFLNENHSSVPIRILQGSIIIVPRIHDEPQAYILQNKEDKRVIFVPIYLDYIS MIGHTDVEYKGDPRKVAITDAERDYLISIVNKHFMREIARSIIAEFSGVRPLCDDDES

[illegible]

/db\_xref="taxon:1309"  
/clone="pMH613"  
complement(<1..28)  
/note="sucrose-6-phosphate hydrolase (scrB)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA26970.2"  
/db\_xref="GI:7272364"  
/translation="MNLQNIRY"  
223..2217  
/note="enzyme scr-II"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA26971.1"  
/db\_xref="GI:153801"  
/translation="MDYSKVAEIVAVGVKDNVLAHAACRLRLVLKDDSKVDOKAL  
DKNADVGFPTQDQVQIIGPDVNFVDEIIKOTGLTEVSDDDLKKAASGKKNP  
IMALKLSDIFVPIIPALVAGGLMALNFLTSEGLFKSLVQOFPPIIKGSSDMIO  
LMSAPFWFLPIVLGISAARFGANQFLGNSIGMIWVAPGAANILGLANAPISKAAT  
IGATGFNFIQGHVTOASTYQVIVPLVAVWLLSILEKFFHRLRPSAVDFTFPLLS  
VIITGLTFIVIGPVKVESDMLTNGIWMLYDFTGLGMVFGALYSPVVMTLGHOSF  
PALEGLISAFQNGTGHGDFIFVTASMANVAQGAATFAIYFLFKDKMKGLSSSGVS  
ALIGITEPALFGVNLAYRPFPCALIGSASAAAIALLOWAVSLGSAAGFLPLSIKA  
SSIPFVVCLEISFAFAVTVYGVKTKADVFAEAFAEAEVEEVEIPEEAASAA  
KAYTDEVLAAPLAGAEVELTSVNDPVFSEAMKGIAPKPSNTVYAPVGVQVIAF  
DHAYGKISDNGAEILIHIGIDITVSMEGKGFQKVOADQIKKGDVIGTFDSDKIAE  
AGLDNTMFVITNDIYASVETILASSTVAVGDSLSLEVRK"  
2227..2252  
misc\_signal  
BASE COUNT 740 a 401 C 513 g 854 t  
ORIGIN  
Query Match 15.1%; Score 230.8; DB 1; Length 2508;  
Best Local Similarity 50.1%; Pred. No. 5.3e-52;  
Matches 779; Conservative 0; Mismatches 717; Indels 58; Gaps 6;

QY 521 ctgtt-----taacagggtggatctctcatcttcctcgaacggcatct 562  
DB 1243 CTCATTTTCAGCTTTCCAAAATGGTACTGGTCATGGTACTGACTTATCTTTGTACAGTTCA 1302  
QY 563 atggctaataatcgcaggtgagcagatgtttggcagtttcttcctcgggaagagtaa 622  
DB 1303 ATGCCCAATTCGCACAGCGCAGCAACCTTTCGAATTTACTCTTAAACAAAAGATAAG 1362  
QY 623 aagctcaaggcccttgagggtgtctccgctgtctcttgggtattacaggagcct 682  
DB 1363 AAGATGAAGAGTCTCTCATCTTCTCAGGTGTGTACAGGCTTCTTGGTATTACAGAACCT 1422  
QY 683 gcatcttcgggtggaacccctgcctgcctgcctgcttcctcatcgggtatcgggacgca 742  
DB 1423 GCTCTTTTGGGGTGAATTTGAAATATCGTTTCCCATTTCTTTGTGCTCTGATTTGGCTCA 1482  
QY 743 gctatcggtgagcgttgattgcaactcttaataataaagcaggtgcttggggcgtgcga 802  
DB 1483 GCTAGTCTGCAGCAATTTGGGGCTTACTTCAAGTTGTAGCGGTTTCTCTTGGTTCGGCA 1542  
QY 803 ggtttcttgggtgtgtttcttattgattgctccagatatggtcattgttcttgggtgtgca 862  
DB 1543 GCTTTTCTTGGTTTCCCTTCTATTAAAGCAAGTCTATTCCATTTATTAGTCTGTGAA 1602  
QY 863 gttgtaccttcttcacatcggtgcgcgcgagattgcttatggcccttacttgggttcgc 922  
DB 1603 TTAATCAGTTTGTCTATAGCTTTTGCAGTAACCTTATGCTATGCTAAGACGAAGCGTGT 1662  
QY 923 cgcaacggcagcattgacagatgcaacgcgtgctccagtgctgcaggaaacacaaa 982  
DB 1663 GATGCTTTTGTGCTGAAGCTGTCTGAAGAAGCTATTGAAGAAGTGCAGAAATTTCCA 1722  
QY 983 gcggaagcagaagcaccgcgagaattttcacaacgattcca---ccatcatccaggcaccct 1039  
DB 1723 GAAGAAGCTGCTTCTGCGCAAAATAAGGCACAACTTACTGATGAAGTCTTTCGGCTCCT 1782  
QY 1040 ttgaccggtgaagcattgcaactgagcagcgcgcagatgcttggcagcggaag 1099  
DB 1783 CTTCGCTGCTGAAGCTGTGAATTAACCTCTGTAAATGACCTGTTTCTTCTAGCGAAGCA 1842  
QY 1100 ctggctcggtggtgctccatgcctcccaacaaaggcaggttagtttctcgggtgagtgga 1159  
DB 1843 ATGGGTGAAGGGGATTTGCTATCAAACTTAGTGTAAATACAGTTTATGACCGGTTGATGGG 1902  
QY 1160 aagattggtggtggtcctccatctggcagtcgttcttcagtcagtcacaaaggtcgaggtat 1219  
DB 1903 ACTGTTCAAAATGCTTTTGATAGTGGCCATGCTTATGGTATTAGTACAGCAATG----- 1957  
QY 1220 ggttccaatgtgatatcttgatgacatgggttcgacacagtaaacacctcaacggcagc 1279  
DB 1958 -----GTGGGAGATTTCTTATTTCATATTGGTATTGATACAGTATCAATGGAAGGTAAA 2010  
QY 1280 cactttaaccgcgtgaagaagcagggcgatgaagtcacaaagcagggggaggtgctgtgtaa 1339  
DB 2011 GGATTTTGAACAAAAAGTTCAAGCAGATCAAAAAATTTAAAAAAGGTGACGTTCTCGAACA 2070  
QY 1340 ttgatattgagtcattgaaggtgcaggttatgagtaaacaccgcgattgttggttcg 1399  
DB 2071 TTTGATAGTACAAAATTCGAGAGCTGGTCTTGGTATTAATACAACTGTTTCATTGTTACT 2130  
QY 1400 aattacaagaaaaccggacccctgtaaacacttacggtttggcgaaattgaagcgggagcc 1459  
DB 2131 AATACAGCTGATTATGCTGAGTTGAAACCTCGCTCTCTTTCAGGAACCTGTTGCTGTAGGT 2190  
QY 1460 ---aacctgctcaacgcgcaagaagaagcgggtgcagcaaacaccataagtt 1510  
DB 2191 GATAGCTTACTTGAAGTTAAAAAATAAGAAATATTATCAGAAAGACCGTAAGGT 2244

RESULT 11  
AE006222/C  
LOCUS  
DEFINITION Pasteurella multocida PM70 section 189 of 204 of the complete  
AE006222 10085 bp DNA linear BCT 08-MAR-2001



|            |   |
|------------|---|
| gene       | complement(9037..9963)  |
| CDS        | /gene="PM1849"  |
|            | complement(9037..9963)  |
|            | /gene="PM1849"  |
|            | /codon_start=1  |
|            | /transl_table=11  |
|            | /product="unknown"  |
|            | /protein_id="AAK03933.1"  |
|            | /db_xref="GI:1272275"   |
|            | /translation="MGKIWLGDVAVDLIPEGENTYLRCAAGAPANVAVGIARIGATT<br>GFTRGVDDPLGRFMLNKLOENVCTQYMILDPKQRTSTVIYDLKEGERSFTTPWPS<br>ADQFLSLDLPHPNMDLHCCSIALINEPSSRSTFAAIIHRAEKGFFSFDPNLRS<br>LWSNPBMKTYPNQVVALDFKSBEELFLTSDTLSDATKLQHYDPDKLIIVTL<br>KLGKATYYLCKQSOLVSSKKALIPVDITAGDAFVSGLLFIISQVKWNQNPTALVEII<br>KQANACGALATTAKGAMSALPNKOQLDPAFLAE" |
| BASE COUNT | 2923 a 1988 c 2212 g 2962 t   |
| ORIGIN     |   |
|            | Query Match 14.3%; Score 218.6; DB 1; Length 10085;   |
|            | Best Local Similarity 54.5%; Pred. No. 1.5e-48;   |
|            | Matches 488; Conservative 0; Mismatches 394; Indels 13; Gaps 2;   |
| Qy 3       | catggcctcgccgcttcgttgcgcttttgccagtgcttggtttgccaccgcaaccaagcg 62   |
| Db 5801    | CATTGC AAAATGC ACTTTTGCTCTTACCTGTATTACTTGGTTTTTCGCCCACTCGAAA 5742   |
| Qy 63      | tctcgccggcaatagattcctggcgccgagtattgtagtgcgatggltcccagact 122  |
| Db 5741    | ATTTGGCGGTAATCCATCTTAGGAGCGGCAC - TTGGCATGCTATTATGTGCCACCGCAC 5683  |
| Qy 123     | tggtaacggctcacgctggccgcaccatggctgcggcgcaaatgccaatgtggtccc 182   |
| Db 5682    | TTTCTGATGGCTGAATATGCACTGCACATCGCTAAAGCAATATTCATATTGGCATA 5623   |
| Qy 183     | tgtttgttttagatgtgtcccaacgcggttacccagggaacogtcttctctgtgtgtg 242  |
| Db 5622    | TC TTTTGGTTT AGAAATTA GAACGTCTCGGCTACCAAGGTACAGTTATCCCGTCTTCTTCG 5563   |
| Qy 243     | tctctgattctggcaacgatcgagaagttctctgcacaagcactcaagggaactgcag 302  |
| Db 5562    | CCTCTGGGCTTTAGCAACCTTTGGAGAAAAACCTACGTAAAGTCGTTCCGTCATTTCTTG 5503   |
| Qy 303     | acttctgatcactccagctgctgacgttgcctgcacggattccttacattccatcccca 362   |
| Db 5502    | ATACTTAATTACCACACTTTTTCATATTATTAAACGGCTTACTCGCTTTTACC GTAA 5443   |
| Qy 363     | ttggccccagaacatcgctgggtggcgatgtctgtggcacacggtctacagggaacttatg 422   |
| Db 5442    | TAGTCTCATCGGCGGTGAAGCGGGCTTTAAATTAGTAGCGGGCTTGACCTGGTTGTATG 5383  |
| Qy 423     | atttcggtgttcagtcggcggtctgtctcttctgttctgttctactcaaccaatcgataca 482   |
| Db 5382    | ACACATTAGGATTCGTGGGGGTGCCAATTTTCGGTACACTTTAGCGACCACTTTGTTATCA 5323  |
| Qy 483     | ctggtctgcaccagtcctt-----ccgcgcaattgagctgagcgtgtttaacc 530   |
| Db 5322    | CAGTATGTCACCAACAATTTATCGCTCTCGNAACACAAATTCCTTGTGAGGTGGCTCGCA 5263   |
| Qy 531     | aggtgtgatccttcatttcgcgaacggcatctatggctaataatgccacaggtgtcgccat 590   |
| Db 5262    | CTGGAGGTACCTTATCTTCCCAATCGCGCGATGTCAAATATTNGCTCAAGGTGCAGCTT 5203  |
| Qy 591     | gtttggcagtgcttcttcctggcgaagagtgaanaagctcaagggccttgacaggtgtcttcag 650  |
| Db 5202    | GTTTGGCGCGCGCTTATGTGATGAAGAATGCCAAGTAAAGGAGATCGCGTACCTTCTG 5143   |
| Qy 651     | gtgtctcgctgttctctgtttattacggagcgtcgactcttcggtgtgaaccttcgctgc 710  |
| Db 5142    | GCATCTCAGCCTGTTAGGGAATTACTGAGCCGCCATGTTTCGGTGTCACTTACGCTATC 5083  |
| Qy 711     | gctggccggtttcttcacgttgatcggtaccgcagctatcggtggcgctttgattgcactct 770  |

|            |      |  |      |
|------------|------|--|------|
| Db         | 5082 | GCTATCCCTTTTCATATCTGCCATGATCGGTGCTGGCGATTCCAGTCCGGTCAATTGCCCTTGT | 5023 |
| Qy         | 771  | ttaataacaaggcaggttcgcttgggcgctcaggtttcttggtggttgcttctatgatg      | 830  |
| Db         | 5022 | TTAATGTTAAGCCATTGCTTTAGCGCTCGGGCTTCGCGGGTATCCCTCAATTAGC          | 4963 |
| Qy         | 831  | ctccagatggtcatgtcttcttggtgtgcagttgttaccattcttcacgcatt            | 885  |
| Db         | 4962 | CGGATAGTTAGCAATGATTTGTGCGGTATGCTTATTTCCGCATCGATTGCATT            | 4908 |
| RESULT     | 12   |  |      |
| VIBCRACK   |      |  |      |
| LOCUS      |      |  |      |
| DEFINITION |      |  |      |
| ACCESSION  |      |  |      |
| VERSION    |      |  |      |
| KEYWORDS   |      |  |      |
| SOURCE     |      |  |      |
| ORGANISM   |      |  |      |
| REFERENCE  |      |  |      |
| AUTHORS    |      |  |      |
| TITLE      |      |  |      |
| JOURNAL    |      |  |      |
| MEDLINE    |      |  |      |
| FEATURES   |      |  |      |
| source     |      |  |      |
| gene       |      |  |      |
| CDS        |      |  |      |
| gene       |      |  |      |
| CDS        |      |  |      |
| gene       |      |  |      |
| CDS        |      |  |      |
| BASE COUNT |      |  |      |
| ORIGIN     |      |  |      |

|                           |        |                  |            |              |
|---------------------------|--------|------------------|------------|--------------|
| Query Match               | 14.1%  | Score 215.2;     | DB 1;      | Length 2955; |
| Best Local Similarity     | 53.9%; | Pred. No. 1e-47; |            |              |
| Matches 493; Conservative | 0;     | Mismatches 408;  | Indels 13; | Gaps 2;      |





|            |   |
|------------|---|
| RESULT     | 14  |
| BSTREAPR   | BSTREAPR  |
| LOCUS      | B.subtilis treA, trep and treR genes.                             |
| DEFINITION | linear BCT 21-NOV-1996  |
| ACCESSION  | 254245  |
| VERSION    | 254245.1  |
| KEYWORDS   | GI:1000450  |
|            | enzyme II-tre; phospho-alpha-(1.1)-glucosidase; phosphotrehalase; |

RESULT 14  
BSTREAPR  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS





```
/product="YfjS"  
/protein_id="BAA233389.1"  
/db_xref="GI:2626811"  
/translation="MKWMSICCAAVLLAGGAAQAAVNEPIMWGFKRSVNHOPDDA  
GKOLSLIKYDAFYLGNTKEKTYILTFDNGYENGYPKVLDVLKKHRYTGTFFVIGH  
FYKDQQLIKRWSDEGHIIIGNSHFHPDLTKTADQIQDELDSVNEEYKIIGKQDNL  
YLRPPGVSEYVLTAKTIGTGIVFWSAFVDWKNQNGKKIAYDHIKMIKQHPGAI  
YLLHTVSRDAAEDDAITDLKQGYTFKSIDDLMEFEKMRPLS"  
complement(1674. .1691)  
complement(1754. .1778)  
complement(1777. .1962)  
/codon_start=1  
/transl_table=11  
/product="YfjT"  
/protein_id="BAA233390.1"  
/db_xref="GI:2626812"  
/translation="MCNAVNNKQQIDLYLNRLDMFMNVIDSLDPEATDVEDIDRLIS  
MLDDLEAKVERKKDWE"  
complement(1960. .1976)  
2090. .2109  
2107. .2571  
/codon_start=1  
/transl_table=11  
/product="YfKA"  
/protein_id="BAA233391.1"  
/db_xref="GI:2626813"  
/translation="MTQNTKLRPITPEFDPWEAYMDVQYQDMQLTNVFTTTLCNM  
RCHKAVGYTLPQKPNALPIDLLKRLKEIPLRSISITGGPMLSLKSVKEYVYPL  
LKVHARGVRTQINSNLTLDIEREWIIPYLDVLHISHNGTVEDEFAEIGA"  
complement(2666. .2701)  
2750. .2768  
2766. .3227  
/codon_start=1  
/transl_table=11  
/product="YfKB"  
/protein_id="BAA233392.1"  
/db_xref="GI:2626814"  
/translation="MYPDSFASALESLSKEMRQATHRLLDIRDENTWMLFGTLPFYA  
CSPDEQKILRLRAKNVTVNRNDPDGRSLNINFDGNIIVTFDGTDPPLGNIQTD  
SLPSAYEKWRETKLAKELSCHCPNRCVLCGNVLVKNYYQDTDFTKRQARS"  
3217. .3230  
3228. .4070  
/codon_start=1  
/transl_table=11  
/product="YfKC"  
/protein_id="BAA233393.1"  
/db_xref="GI:2626815"  
/translation="MEKELTEIFONKIVDILLVAVILWIGVFINRLVQLFFKRTD  
FIEEKKEITSLVRSTQITATIGFIVISLFDHDFGKILAGAVAGIVIGFGAQS  
LKDVLAGVLIYEROLHKGDDYVNVNLFNGVTEEIGLSLQIREWSGKLLITSNGEV  
RQIENYIDFMRTESFLISFKEDPDRTVSVLEACDMLNEELRDSLKRFECNPPEP  
FQIHGTALNKINRGVEFTVKGMDKDDYFASLAVRRVLVRLQYQNNVQMLEAVRI  
ERTQ"  
complement(4074. .4106)  
complement(4109. .4903)  
/codon_start=1  
/transl_table=11  
/product="YfKD"  
/protein_id="BAA233394.1"  
/db_xref="GI:2626816"  
/translation="MMKLPFHTLVLVLFSEFGVQPIHAKQKQKVPNVSASIKENT  
YPNASODQPMLOPSKIAKELLDHSEVKIENPHILMKNESINSGTPLAYGIRATIFLG  
KWALGYSETNANWEYKIKINTNRADNRGKGKETAEMHYAQEQQYRVKGGILAKRVPAE  
DYKSMQMKAMKTNLPAPFETVIGAGTRKDQIYKVAPKKIGYLHAYAPAVNEKGVV  
YGEVYLVLKGNKRKLVKNVTSGQIGAWIPVDHVTFFGQLSSLPR"  
complement(4901. .4913)  
complement(4974. .6029)  
/codon_start=1  
/transl_table=11  
/product="YfKE"  
/protein_id="BAA233395.1"  
/db_xref="GI:2626817"  
/translation="MNRIFILVAGVPLSVIGSLMHPVSAVLFAVYCVTIILASYM
```

```
GRATESLSIAGPRIGGLLNATGNAVELIISFALKEGTGTIVLASLTGSLVGLNLL  
VAGLSFFGLKTKRQEFNIHDRHNSGGLLFAIIVAFVPEVFSVGMGNKLSLSI  
GISIMILLYVAALYFKLVTHRGVYQNNAAQTEEPPEWSGKVATIVLFAATIVVA  
YISENVHVFHSHVABQFWSSELFIVIIAIVGNAEAHASAIIMAKNKNMDIAVEIAV  
GSTLTAMVAPVVLVICSIFPPTSMPLVFTLPELVAMVSAVLLMIALISNDGDSNWFE  
ATLLAQYVMAIGFFLL"  
complement(6027. .6046)  
6180. .6199  
6197. .7372  
/codon_start=1  
/transl_table=11  
/product="YfKF"  
/protein_id="BAA23396.1"  
/db_xref="GI:2626818"  
/translation="MSRFHFLLVLLVLSISGFSQGMLLPVISIIFTNGSAAINGLH  
ATGLYIGVLLASFPMEAPLRKLGFKPLIVMGGSIVLSLFGFTWLSQSVWVWFLRLFI  
GIGDHMLHSTQVWTVSMSSKQNRNLSIYGLSGFGAAGFPFVPLVKLSPLSFPFI  
VSGCISLFAWLVFFLQNAYPETSPHETKSDNSFRFYQAMLFGWAFMFTFGYGFLE  
TALNGSFPVYALRGISVDVAIILPAFAIGSIIFQFPLGIDSGYGRNVLLVILT  
GALCFFIAGVFPSPYVIGGCFIAGMAGVSTFTLGISYMTDLPLPPLLPAAGNLICGT  
FSLGSILGPVAGGWYQTESANLFFITLTSSVWLALVLGPKMSWSPAETYSSES"  
complement(7376. .7408)  
7396. .7411  
7409. .7546  
/codon_start=1  
/transl_table=11  
/product="YfKG"  
/protein_id="BAA23397.1"  
/db_xref="GI:2626819"  
/translation="MLWAVLSVGASACFMARFAFVLVFMASAFRMLGHHDIPPTS  
F"  
complement(7571. .8398)  
/codon_start=1  
/transl_table=11  
/product="YfKH"  
/protein_id="BAA23398.1"  
/db_xref="GI:2626820"  
/translation="MSFLKELFSRYTLHEGQSKSAELAYFFLLSLFPFLIFMLTLTAY  
LPLSTDVLGVIQYAPASAMLSVESITHTLNNRGGLLSFGIIAALWSASNGMAI  
VRLSHAYDVEENRSFIIIVRLTSIFLTIAMVFTILVALLLPVGREIGRLASDFVCA  
DPLFSWAAIRWGVSPVLVLLIVFSALYVIAPNKKLSLRFVWPGAVFATIGWILVSLF  
SFVSTFANISATYSGIGGIIVLMIWFIUSGILIIIGGEINALLHKKRKLDPDENPYH"  
complement(8396. .8415)  
complement(8422. .8742)  
/codon_start=1  
/transl_table=11  
/product="YfKI"  
/protein_id="BAA23399.1"  
/db_xref="GI:2626821"  
/translation="MNERMLKGIIFLGAAGAALSLLHKKPTRQACGMRLWLTCKHKLS  
LYKSNPELLKNTVITKVDKAKKLARTLSKEVDVFNQOVKELKKTTPQVMELVQETKEH  
FSKK"  
complement(8740. .8756)  
complement(8760. .9230)  
/codon_start=1  
/transl_table=11  
/product="YfKJ"  
/protein_id="BAA23400.1"  
/db_xref="GI:2626822"  
/translation="MISLVFVLGNTRCRSPMAEATFRDLAAKGLSGKTKADSAGTGG  
WHIGNPHGTOEILRRREGISFQDMLARQVSEQDDDFDYIIMADENIGSLRSAGF  
KNTSHKRLLDVVEDSLADVPDPPYTGNEFEEVQCLIKTGCEOLLASIOKEKOL"  
complement(9228. .9247)  
9418. .9435  
9433. .9648  
/codon_start=1  
/transl_table=11  
/product="YfKK"  
/protein_id="BAA23401.1"  
/db_xref="GI:2626823"  
/translation="MSSPNTETLTQMIEELSOKLMLNVGVKAEDEFSDEKIEDLTYL  
HRVMKKSFSFSEMQIAQELASLRK"  
9727. .9743
```

|     |  |  |                 |                    |           |               |
|-----|--|--|-----------------|--------------------|-----------|---------------|
| CDS | 9741..10931  | Query Match  | 13.2%;          | Score 201.8;       | DB 1;     | Length 22197; |
|     | /codon_start=1   | Best Local Similarity  | 52.9%;          | Pred. No. 6.7e-44; |           |               |
|     | /transl_table=11   | Matches 482;   | Conservative 0; | Mismatches 422;    | Indels 7; | Gaps 2;       |
|     | /product="Yfkl"  |  |                 |                    |           |               |
|     | /protein_id="BAA23402.1"                                 |  |                 |                    |           |               |
|     | /db_xref="GI:2626824"                                    |  |                 |                    |           |               |
|     | SYVIAAGFGKLLTYGISGVLSDKIGRKPLVVASAGINAVFLVGLPSPVELAFVFA  |  |                 |                    |           |               |
|     | LLAGVANSAMDAGTYPALTELPSPASGSANLVKAFMSVGAALLPLLTFLADHSMFY |  |                 |                    |           |               |
|     | GFAFYLPAAVYLLNLIYLSLTFPKKKHKPTNSGQOESPVLSEPFVFKEGTALLIG  |  |                 |                    |           |               |
| QY  | 1  | ctcatggcaatcgcgcgcttcgcttcttgccagtgattggttggtttccaccgcaaccaag    | 60              |                    |           |               |
| DB  | 20384  | CTGATGTGACAGCACGGCCCTTTACGCTTTTGCCCTGCGTTAATCGGATGGTCTGCCGTC     | 20325           |                    |           |               |
| QY  | 61   | cgtttcgcgcgcaatgagttcctggcgccgcgctattggtatggtgagatggtttcccgag    | 120             |                    |           |               |
| DB  | 20324  | CGGTTGCGGGCGCAATCCGGCTTCTCGGCATTTGCG-TCGGGGTTATGCTCGTGCACTCTGA   | 20266           |                    |           |               |
| QY  | 121  | cttggtgaacggtctacgctggtgcgcgccaccatggctgcggcgcaaatgccaatgtggtc   | 180             |                    |           |               |
| DB  | 20265  | TTTGCTGAATGCTGGGGATATGGCGCGGCAGACAAACGCGAGAGATCCCGGTATGGAA       | 20206           |                    |           |               |
| QY  | 181  | ctctgttggttagatggttgcgaagccggtttaccaggcagcgtgcttccctgtgctggt     | 240             |                    |           |               |
| DB  | 20205  | TCTATTTCGGCCTAGAGGTGCAGAAAGTCGGCTATCAGGCGCAGGTGCTCCCAATTTTGCT    | 20146           |                    |           |               |
| QY  | 241  | ggtttcttggtattctggaacatcgagaagttcctgcacaagcgactcaagggaactgc      | 300             |                    |           |               |
| DB  | 20145  | TGCTTCTTATATGCTGGCGAAGATTGAGGTATTTTAAACAAACGCACACCTGAGGGCAT      | 20086           |                    |           |               |
| QY  | 301  | agacttcctgatacctccagtgctgacgttgctgcaccggattccttacattcctcgc       | 360             |                    |           |               |
| DB  | 20085  | ACAGCTGCTCGTTGTCGCACCAATTACGCTTCTCTTAACAGGATTTGCCCTCTTTATCAT     | 20026           |                    |           |               |
| QY  | 361  | cattggccacgaatgcgctggtggtggcgatgctggtgcacacggtctcacagggaacttta   | 420             |                    |           |               |
| DB  | 20025  | TATCGCCCGGATTACATTTTGCATCGGTAACTACTGACTTCAGGGCTTATCTCGGTGTT      | 19966           |                    |           |               |
| QY  | 421  | tgatttcggtggtccagtcggtgctctcttcggtctggtggtctactcacaatgctcat      | 480             |                    |           |               |
| DB  | 19965  | TGGTCTTTTTCGGCGCTGGCGGTCTGTTATATGGCGGTTTCTACTACGCGCTCGTGAT       | 19906           |                    |           |               |
| QY  | 481  | cactggtctgcaccagtccttcccgccaattgagctggagct-----gtttaaccaggg      | 534             |                    |           |               |
| DB  | 19905  | TACCGGTATGCATCACACGCTTTCTTGCAGCTCGACCTGCAGCTCATCGGTTCAAAGCTCGG   | 19846           |                    |           |               |
| QY  | 535  | tggatccttcaatcttgcacaacggcatctatggtctaataatgcgccaggggtgcgcatgttt | 594             |                    |           |               |
| DB  | 19845  | CGGCACCTTTTATGCGCGATGCTGGCGGCTTCTTAATATCGCACAAAGGTTTCGGCGGCTCT   | 19786           |                    |           |               |
| QY  | 595  | ggcagtggttcttcctggggaagagtgaaaagctcaaggcccttgcaagtgcttcaaggtgt   | 654             |                    |           |               |
| DB  | 19785  | GGCCATGATGTTTATCGTGAAGGATGAGAAACAAAAGGCCCTCCCTCACATCAGGGAT       | 19726           |                    |           |               |
| QY  | 655  | ctccgctgcttctggtattacggagcgtcgatcttcggtgtgaaccttcgctgcgctg       | 714             |                    |           |               |
| DB  | 19725  | TTCTGCTTATCTTGGCATTTACAGAGCCTGCCATCTTTGGGGTGAATCTCGGATACAGATT    | 19666           |                    |           |               |
| QY  | 715  | gcggttcttcacgtgtatcgtgtaccgcagctatcgggtggcgcttttgattgcaactcttta  | 774             |                    |           |               |
| DB  | 19665  | CCGCTTTATCATTTGCAATGGTGCAGCTCTGGCCTGGCGGGAATGATATTTCTTCTCAAGG    | 19606           |                    |           |               |
| QY  | 775  | tataaggcagttgcgttgggcgctgcaggttctcttggtggtgtgttcttatgtatgctcc    | 834             |                    |           |               |
| DB  | 19605  | TCGTGTGGCAAGCTCCGTGGTGTGCGGCGGTGCGGGGATTTTCTCAATCATCAGCCA        | 19546           |                    |           |               |
| QY  | 835  | agatatggtcatgttcttggtgtgtgagttgttacccttcttcacgtcgttcggcgagc      | 894             |                    |           |               |

Db 19545 GTACTGGGCGCGTTTTGCAATCGGGATGCGTATTGATGTCGCCGTTTTCGCCGAAC 19486

QY 895 gattgcttatg 905

Db 19485 ATACGCGTATG 19475

Search completed: September 25, 2002, 07:30:47

JOB time: 5098 sec

**THIS PAGE BLANK (USPTO)**